DETECTING SQUAREFREE NUMBERS

ANDREW R. BOOKER, GHAITH A. HIARY, AND JON P. KEATING

Abstract. We present an algorithm, based on the explicit formula for L-functions and conditional on GRH, for proving that a given integer is squarefree with little or no knowledge of its factorization. We analyze the algorithm both theoretically and practically, and use it to prove that several RSA challenge numbers are not squarefull.

1. Introduction

Let k be a finite field and f a non-zero element of $k[x]$. Then it is well known that f is squarefree if and only if $\gcd(f, f') = 1$, and the latter condition may be checked quickly (in deterministic polynomial time) by the Euclidean algorithm. It is a long-standing question in algorithmic number theory whether there is a correspondingly simple procedure to test if a given integer is squarefree; in particular, can one determine whether $N \in \mathbb{Z}$ is squarefree more rapidly than by factoring it?

In this paper, we describe an algorithm, conditional on the Generalized Riemann Hypothesis (GRH), for proving an integer squarefree with little or no knowledge of its factorization, and analyze the complexity of the algorithm both theoretically and practically. In particular, we present some heuristic evidence based on random matrix theory and other probabilistic calculations that our algorithm runs in deterministic subexponential time $O\big(\exp[(\log N)^{2/3+o(1)}]\big)$. Although this is poorer than the performance expected of the current best known factoring algorithms, our method is able to give partial results that one does not obtain from a failed attempt at factoring. In particular, we show the following (see §3.2).

Theorem 1.1. *Assume GRH for quadratic Dirichlet L-functions. Then the RSA challenge numbers* `RSA-210`*,* `RSA-220`*,* `RSA-230` *and* `RSA-232` *are not squarefull, i.e. each has at least one prime factor of multiplicity* 1*.*

The challenge numbers mentioned in the theorem, ranging in size from 210 to 232 digits, are significant because they are the smallest that have yet to be factored.[1] Certainly the technology to factor them exists (in fact the comparably sized[2] `RSA-704` and `RSA-768` were successfully factored in 2012 and 2009, respectively), but it remains prohibitively expensive to perform such factorizations routinely. In contrast, the proof of Theorem 1.1 for `RSA-210`

A. R. B. was supported by EPSRC Fellowship EP/H005188/1. J. P. K. was sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA8655-10-1-3088. The US Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon. J. P. K. and G. A. H. also gratefully acknowledge support from the Leverhulme Trust.

[1] `RSA-210` was factored in September 2013, after this paper was submitted but before publication.

[2] `RSA-704` and `RSA-768` are named for their sizes in binary; in decimal they have 212 and 232 digits, respectively.

could be carried out with a desktop PC in a few months. To our knowledge, Theorem 1.1 is the first statement of its kind to be proven without exhibiting any factors of the number in question.

Acknowledgements. We thank Paul Bourgade, Peter Sarnak and Akshay Venkatesh for helpful conversations.

1.1. **Background.** We begin with some background on the problem of squarefree testing, before describing our main algorithm in §2. Given an integer $N > 1$, we first note that if N has no prime factors $\leq \sqrt[3]{N}$ then it is squarefree if and only if it is not a perfect square. Thus, since it is easy to detect squares, in order to prove a number squarefree it suffices to find all of its prime factors up to the cube root. On the other hand, the Pollard–Strassen algorithm [25, 30] finds all prime factors of N up to a given bound B in time $O_\varepsilon\big(N^\varepsilon\sqrt{B}\big)$. This immediately yields an algorithm for squarefree testing in time $O_\varepsilon\big(N^{1/6+\varepsilon}\big)$. We remark that with some modifications to the Pollard–Strassen algorithm, along the lines of [5] but specific to this problem, one can improve the running time very slightly to $O\big(N^{\frac{1}{6}-\frac{c}{\log\log N}}\big)$ for some $c > 0$.

Although Pollard–Strassen is often regarded as a purely theoretical result, with modern computers it is possible to implement it and realize some improvement in speed over trial division. However, the gains do not occur until B is of size 10^9 at least. As a result, even the modified algorithm mentioned above is only practical for N up to 10^{70} or so. On the other hand, the Quadratic Sieve algorithm running on a PC will, in practice, almost surely factor a given $N \leq 10^{70}$ within a few minutes; thus, at least with present algorithms and technology, it is always better to try to factor the given integer.

1.2. **Fundamental discriminants.** Our approach rests on a way of characterizing the squarefree integers that does not directly refer to their factorization. Precisely, if $d \in \mathbb{Z}$, $d \equiv 1 \pmod 4$, then d is squarefree if and only if it is a fundamental discriminant. (Note that if $N \in \mathbb{Z}$ is odd then $d = (-1)^{\frac{N-1}{2}}N$ satisfies $d \equiv 1 \pmod 4$, so this restriction entails no loss of generality.) The advantage of this criterion is that whether or not a given discriminant d is fundamental can be detected from values of the quadratic character $\chi_d(n) = \left(\frac{d}{n}\right)$, where $\left(\frac{\cdot}{\cdot}\right)$ denotes the Kronecker symbol. In turn, $\chi_d(n)$ is easy to compute for a given n, thanks to quadratic reciprocity; in particular, if n is a prime then the Kronecker symbol $\left(\frac{d}{n}\right)$ reduces to the Legendre symbol, which can be evaluated, e.g., by Euler's criterion.

Let $\mathcal{F}$ denote the set of fundamental discriminants. To see how one might use the above to prove quickly that a given d is squarefree, note first that we have in general that $d = \Delta\ell^2$, where $\Delta \in \mathcal{F}$ and $\ell \in \mathbb{Z}_{>0}$. Here $|\Delta|$ is an invariant of the character χ_d (its conductor), which we aim to show equals, or is as least close to, $|d|$. By testing whether d is a square, we may assume without loss of generality that $\Delta \neq 1$.

For any $x > 0$, consider the series

$$S_\Delta(x) = \frac{1}{\sqrt{x}}\sum_{n=1}^{\infty}\chi_\Delta(n)\left(\frac{n}{x}\right)^{(1-\chi_\Delta(-1))/2}e^{-\pi(n/x)^2}, \tag{1}$$

which is essentially the twisted θ-function. Note here that we may calculate $\chi_\Delta(n)$ for any given n, even without knowledge of Δ; in fact, we have $\chi_\Delta(n) = \chi_d(n)$ unless n has a common

factor with ℓ. We may assume without loss of generality that we never come across such an n, since otherwise we will have found a square factor of d, answering the original question.

If one thinks of the character values $\chi_\Delta(n)$ as "random" ± 1 then, thanks to the decay of the Gaussian, the series in (1) is the result of a random walk of length about x, which will typically have size on the order of $\sqrt{x}$; thus, one might expect $S_\Delta(x)$ to oscillate, without growing very large or decaying, as $x \to \infty$. This turns out to be an accurate description for x up to $\sqrt{|\Delta|}$, but for larger x, $S_\Delta(x)$ is constrained by the symmetry

$$S_\Delta(x) = S_\Delta(|\Delta|/x), \tag{2}$$

following from the Poisson summation formula (see [6, pp. 13, 68, 70]).

The point of symmetry of (2) gives an indication of $|\Delta|$, and thus we can rule out small values of $|\Delta|$ essentially by drawing the graph of $S_\Delta(x)$. More precisely, for any given $B > 0$, one can decide whether or not $|\Delta| \le B$ in time[3] $O_\varepsilon\big(N^\varepsilon \sqrt{B}\big)$, which matches the running time of Pollard–Strassen for the same task. Moreover, if one could find a method of computing the θ-function $S_\Delta(x)$ substantially more quickly than by direct in-order summation, say in time $O_\varepsilon\big(N^\varepsilon x^{1-\delta}\big)$ for some $\delta \in (0,1)$, then this improves to $O_\varepsilon\big(N^\varepsilon B^{\frac{1}{2}(1-\delta)}\big)$; in particular, taking $B = N^{\frac{1}{3-2\delta}}$ and falling back on Pollard–Strassen to rule out $\ell \le \sqrt{N/B}$, we would get an algorithm to certify N squarefree in time $O_\varepsilon\big(N^{\frac{1}{6}(1-\frac{\delta}{3-2\delta})+\varepsilon}\big)$.

2. The explicit formula

Our main interest, however, is in algorithms that work in subexponential time. This is difficult to attain in the above approach because we used sums over integers n. It is well-understood in problems of this type that one can do better by considering sums over primes, at the expense of having to assume GRH.

To be precise, let $L(s,\chi_\Delta) = \sum_{n=1}^\infty \chi_\Delta(n) n^{-s}$ be the Dirichlet L-function corresponding to $\Delta \ne 1$. Assuming GRH, the non-trivial zeros of $L(s,\chi_\Delta)$ may be written as $\frac{1}{2} \pm i\gamma_j(\Delta)$, $j = 1,2,3,\ldots$, where $0 \le \gamma_1(\Delta) \le \gamma_2(\Delta) \le \ldots$, and each ordinate is repeated with the appropriate multiplicity.[4] Further, let $g : [0,\infty) \to \mathbb{C}$ be a test function which is continuous of compact support, piecewise smooth, and has cosine transform $h(t) = 2\int_0^\infty g(x)\cos(tx)\,dx$. Then the "explicit formula" for $L(s,\chi_\Delta)$ reads

$$\begin{aligned} g(0)\log|\Delta| &= 2\sum_{j=1}^\infty h\big(\gamma_j(\Delta)\big) + 2\sum_{n=1}^\infty \frac{\Lambda(n)\chi_\Delta(n)}{\sqrt{n}} g(\log n) \\ &\quad + g(0)\log(8\pi e^\gamma) - \int_0^\infty \frac{g(0)-g(x)}{2\sinh(x/2)}\,dx + \chi_\Delta(-1)\int_0^\infty \frac{g(x)}{2\cosh(x/2)}\,dx, \end{aligned} \tag{3}$$

where Λ is the von Mangoldt function.

If not for the sum over zeros $\gamma_j(\Delta)$, this would be exactly what we seek, i.e. a formula for the conductor $|\Delta|$ in terms of character values. Without knowledge of the zeros, we do not

[3] We omit the proof of this, but the main point is the fact that $S_\Delta(e^x)$ is the Fourier transform of the complete L-function $\Lambda(\frac{1}{2}+it,\chi_\Delta)$, so it is essentially band-limited.

[4] If $L(s,\chi_\Delta)$ has a zero at $s = \frac{1}{2}$ of multiplicity m, then m is necessarily even, and we take $\frac{m}{2}$ copies of this zero, i.e. $\gamma_j(\Delta) = 0$ for $j \le \frac{m}{2}$ and $\gamma_{\frac{m}{2}+1}(\Delta) > 0$.

get such an exact identity, but we can at least get an inequality in one direction if the test function is chosen so that h is non-negative, i.e.

$$\begin{aligned}\log|\Delta| \geq 2\sum_{n=1}^{\infty}\frac{\Lambda(n)\chi_\Delta(n)}{\sqrt{n}}g(\log n) + \log(8\pi e^\gamma) \\ - \int_0^\infty \frac{1-g(x)}{2\sinh(x/2)}\,dx + \chi_\Delta(-1)\int_0^\infty \frac{g(x)}{2\cosh(x/2)}\,dx,\end{aligned} \tag{4}$$

for any $g:[0,\infty)\to\mathbb{R}$ which is continuous of compact support, satisfies $g(0)=1$, and has non-negative cosine transform.[5]

In the next few subsections we explore some strategies for exploiting (4) to prove that our given d is squarefree. The proofs of Propositions 2.1–2.3 below are given in the appendix.

2.1. **Varying the test function.** Our first, and simplest, strategy is to search for a test function such that the right-hand side of (4) is close to $\log|\Delta|$. Naturally, we pay a price for ignoring the zero sum $Z=\sum_{j=1}^{\infty}h(\gamma_j(\Delta))$, in that our estimate for $|\Delta|$ is a factor of e^{2Z} too small. We can still use this to prove that d is squarefree by ruling out values of $\ell\leq e^Z$ using Pollard–Strassen or otherwise, but this takes exponential time $\gg e^{Z/2}$ in the size of Z.

On the other hand, note that the sum over prime powers in (4) is exponentially long, i.e. if g has support $[0,X]$ then we need to compute the right-hand side of (4) for n up to e^X. Thus, we would like X not to be very large. However, if we choose X too small then, by the uncertainty principle, the cosine transform h will be relatively "wide", so that the zero sum will typically be large.

Our first result shows that there is an optimal choice of test function for each fixed X, and thus an optimal tradeoff between these two exponential penalties.

Proposition 2.1. *Let $\mathcal{C}(X)$ be the class of functions $g:[0,\infty)\to\mathbb{R}$ that are continuous, supported on $[0,X]$, have non-negative cosine transform, and satisfy $g(0)=1$. For $g\in\mathcal{C}(X)$, let $l(g)$ denote the right-hand side of (4). Then for every $X>0$ there exists $g_X\in\mathcal{C}(X)$ such that $l(g_X)\geq l(g)$ for all $g\in\mathcal{C}(X)$.*

We remark further that if $g\in\mathcal{C}(X)$ then its cosine transform h is band-limited, and so, by Jensen's formula, h has at most $O_X(T)$ zeros in the interval $[-T,T]$ for large T (see [19, p. 16]). Since it is known that $L(s,\chi_\Delta)$ has $\gg T\log T$ distinct zeros with imaginary part in $[-T,T]$, under GRH the zero sum in (3) cannot vanish, so that (4) is a strict inequality for any fixed X, i.e. $\log|\Delta|>l(g_X)$. However, it is easy to see that $l(g_X)$ tends continuously and monotonically to $\log|\Delta|$ as $X\to\infty$.

Although Prop. 2.1 is an existence result only, one can try to solve for the optimal test function g_X by approximating $\mathcal{C}(X)$ using a sufficiently rich, finite-dimensional space of functions. For instance, let M be a non-negative integer, and consider step functions f of

[5] A function g satisfying these conditions need not be piecewise smooth, and in fact $\int_0^\infty \frac{1-g(x)}{2\sinh(x/2)}\,dx$ may be divergent. However, since g has non-negative cosine transform, $\frac{1-g(x)}{2\sinh(x/2)}$ is non-negative, so we may interpret the right-hand side of (4) as $-\infty$ whenever the integral diverges. With that convention, a standard approximation argument shows that (4) holds for all g as indicated.

the form

$$f(x) = \sum_{n=-M}^{M} a_n \mathbf{1}_{(-1/2,1/2)}\left(\frac{2M+1}{X}x - n\right) \quad \text{for } x \in \mathbb{R}, \tag{5}$$

for arbitrary real coefficients a_n. If we take g to be the autocorrelation of f, i.e. $g(x) = \int_{\mathbb{R}} f(y)f(x+y)\,dy$, then g has cosine transform $\left|\hat{f}(t)\right|^2 \geq 0$, the right-hand side of (4) is a quadratic form in the a_n, and the condition $g(0) = 1$ amounts to an L^2-normalization. Thus, we can find the optimal lower bound for this family of test functions by computing the matrix of the form and finding its largest eigenvalue.[6] It is not hard to see that this family comes arbitrarily close to the optimal g_X as $M \to \infty$, although it may be the case that g_X is highly oscillatory, meaning that we would need to take M very large before finding a close approximation to it.

2.2. **Twisting.** A second strategy, which performs well in practice, is to "twist" our given quadratic character χ_d by other characters χ_q, and look for a q for which the lower bound in (4) is favorable. This is related to the first strategy since, by Fourier analysis, varying the test function amounts to considering combinations of the twists by n^{it} for various t. Twists by quadratic characters have the added advantage of zero repulsion around the central point, as we explain in detail in §3.1.

In other words, if we run out of luck with our given value of d then we can multiply it by $q \in \mathcal{F}$ relatively prime to d and ask if the product is a fundamental discriminant. This operation also introduces a penalty, since (4) becomes a lower bound for $\log|q\Delta|$, so we have to subtract $\log|q|$:

$$\begin{aligned}\log|\Delta| \geq -\log|q| + 2\sum_{n=1}^{\infty}\frac{\Lambda(n)\chi_{q\Delta}(n)}{\sqrt{n}}g(\log n) + \log(8\pi e^{\gamma}) \\ - \int_0^{\infty}\frac{1-g(x)}{2\sinh(x/2)}\,dx + \chi_{q\Delta}(-1)\int_0^{\infty}\frac{g(x)}{2\cosh(x/2)}\,dx.\end{aligned} \tag{6}$$

What we gain by this strategy is the hope of finding a twist $\chi_{q\Delta}$ such that the low-lying zeros of $L(s,\chi_{q\Delta})$ are unusually sparse, so that the zero sum $\sum_{j=1}^{\infty} h(\gamma_j(q\Delta))$ can be made small even with a relatively simple choice of g. For instance, we might hope that $L(s,\chi_{q\Delta})$ has a large zero gap around the central point. In that case, we have the following.

Proposition 2.2. *Suppose that $L(s,\chi_{q\Delta})$ satisfies GRH and has no non-trivial zeros with imaginary part in $(-\delta,\delta)$. Set $X = 2\delta^{-1}(A + \log\log|q\Delta|)$ for some $A \geq 0$. Then there is an explicit $g \in \mathcal{C}(X)$ whose cosine transform h satisfies*

$$\sum_{j=1}^{\infty} h(\gamma_j(q\Delta)) \ll \frac{e^{-A}X}{(\log\log|q\Delta|)^{3/2}}, \tag{7}$$

with an absolute and effective implied constant.

[6] If A is the matrix associated with the quadratic form and $c := (2M+1)/X$, then $\max_{|a|^2=c} a^t A a = c\lambda_1$, where λ_1 is the largest eigenvalue of A, $a := (a_{-M},\dots,a_M)$, and the condition $|a|^2 = c$ is equivalent to $g(0) = 1$.

In other words, there is a test function g with support of size inversely proportional to the size of the zero gap for which the zero sum is relatively small. Thus, ruling out small values of ℓ to complete the proof that d is squarefree is fast compared to evaluating the explicit formula.[7]

Although it is difficult to ascertain directly for a given q whether $L(s, \chi_{q\Delta})$ has a large zero gap, we can simply try computing the lower bound (6) using the test function given by Prop. 2.2 for a particular desired value of δ. We may repeat this procedure for many q until we find one which is good enough, and then use the quadratic form approach with a relatively small matrix to refine the choice of test function.

The crucial question is thus how large of a zero gap can one expect to find by searching through various q. On average, one expects the first zero gap around the central point to be about $2\pi/\log|q\Delta|$,[8] which is of no use in Prop. 2.2. On the other hand, if we found q of modest size for which the first zero gap was on the order of $1/\sqrt{\log|q\Delta|}$, say, then we would have a fast algorithm for proving that d is squarefree.[9]

To make this more precise, anticipating a subexponential running time on the order of $\exp\big((\log|\Delta|)^\theta\big)$, for $\theta > 0$ we define

$$M_\Delta(\theta) = \max\Big\{\gamma_1(q\Delta) : q \in \mathcal{F}, (q,\Delta) = 1, |q| \le \exp\big((\log|\Delta|)^\theta\big)\Big\},$$

$$\eta_\Delta(\theta) = -\frac{\log M_\Delta(\theta)}{\log\log|\Delta|}, \quad \eta_\infty(\theta) = \limsup_{\substack{\Delta\in\mathcal{F}\\ |\Delta|\to\infty}} \eta_\Delta(\theta), \quad \theta^* = \inf\{\theta > 0 : \eta_\infty(\theta) \le \theta\}.$$

Thus, η_Δ is a logarithmic measure of the largest gap size that we encounter among the twists by $q \in \mathcal{F}$ with $|q| \le \exp\big((\log|\Delta|)^\theta\big)$, with $\eta_\Delta = 1$ corresponding to an average gap, and $\eta_\Delta < 1$ corresponding to larger gaps. Since, *a priori*, we have no information about our given discriminant, we take the worst case, η_∞, over all large values of Δ. Finally, θ^* measures the point at which the size of the twisting set matches the expected length of the prime sum that we need to evaluate in Prop. 2.2. Combining this with a brute-force search strategy, we obtain the following.

Proposition 2.3. *Assume GRH for quadratic Dirichlet L-functions. There is an algorithm that takes as input a positive integer N and outputs either a non-trivial square factor of N or a proof that N is squarefree. If N is squarefree then the algorithm runs in time $O\big(\exp[(\log N)^{\theta^*+o(1)}]\big)$.*

Note that the assumption of GRH in the proposition applies to the certificates generated by the algorithm as well as its running time analysis.

---

[7] In fact, as a by-product of evaluating the explicit formula, we will test d for divisibility by all primes $p < e^X$. Thus, if $\sum_j h(\gamma_j(q\Delta)) < X$ then no additional work is necessary to prove that d is squarefree.

[8] More precisely, the Random Matrix model for the family of L-functions in question suggests that the mean of the first zero gap should be this quantity multiplied by a constant whose value is approximately 0.78, see [14, 26]

[9] If there is a constant $\varepsilon > 0$ such that one can always find a $\gamma_1(q\Delta) \ge (\log\log|\Delta|)^{1+\varepsilon}/\log|\Delta|$, then Prop. 2.2 already allows one to certify that an integer is squarefree in subexponential time (on the GRH). It would be interesting to see if one could push this line of thought to an improvement of the $O(N^{1/6+o(1)})$ time bound of Pollard–Strassen, but we do not do so here.

2.3. **Examples.** In Figure 1, we give a basic illustration of the favorable situation of a large gap around the central point. We chose $L(s, \chi_d)$, where $d = 1548889$ is a fundamental discriminant, because it has a gap size ≈ 1.747424 there, which is about 4.5 times the average $0.78 \times 2\pi/\log(d/(2\pi))$. Therefore, we expect the lower bound (4) to be quite good even with a simple choice of g. For instance, if $M = 0$ and $a_0 = 1/\sqrt{X}$ in (5), we obtain $g(x) = \max(0, 1 - |x|/X)$ and $h(t) = X \sin^2(Xt/2)/(Xt/2)^2$. Choosing $X = 7/2$, we have $2\sum_{j\geq 1} h(\gamma_j(d)) \approx 6.73$ (by computing the zeros explicitly using `lcalc`), and so the lower bound (4) would be $\log d - 6.73 \approx 7.5$. This would have sufficed to prove that d is squarefree, since in computing the prime sum of the explicit formula we would have checked that d has no factor $\leq e^{7/2}$, and so certainly no factor $\leq e^{6.73/2}$. In particular, (4) allows one to certify that d is squarefree from the primes $\leq e^{7/2} \approx 33$ only, which is is already better than trial division. Thus, our strategy can lead to a gain even for small d.

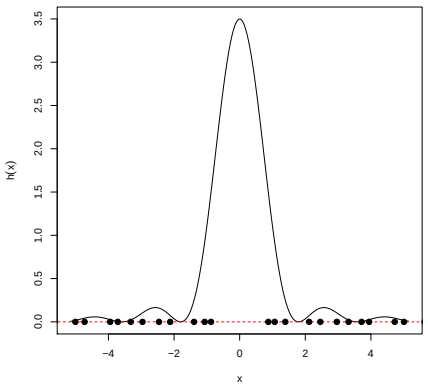

FIGURE 1. Large zero gap around the central point of $L(s, \chi_{1548889})$, together with the test function on the zero side $h(t) = 8\sin^2(7t/4)/(7t^2)$ resulting from $M = 0$ and $X = 7/2$.

The behavior of the lower bound as X increases is worth noting. We illustrate it for $L(s, \chi_d)$, Figure 2 (left plot), using the same simple choice of g as before. The overall shape of the plot is typical for the case of a large gap, in that there is an initial (good) region where the lower bound increases steeply, followed by an inevitable, unless $L(1/2, \chi_d) = 0$, region of small oscillations. If the gap about the center is not particularly large, however, then the initial good region will be much smaller. This is illustrated in Figure 2 (right plot) using the L-function of a randomly chosen fundamental discriminant, $L(s, \chi_{2000005})$, which has an average-sized gap of ≈ 0.515984 about the center. Notice that there is a wide good region later on in the plot, but it comes in too late to be useful in our algorithm. The main point is that the absence of zeros near $s = 1/2$ allows the sum over prime powers to capture the bulk of the r.h.s. of the explicit formula (3) with a smaller choice of X (i.e. a more compactly supported g, and slower decay for h on the zeros sum).

3. Complexity

By computing the 1-level density of the family of twists by χ_q, $q \in \mathcal{F}$, one can see that $\eta_\infty(\theta) \leq 1$ for $\theta > 1$, so that $\theta^* \in [0, 1]$. However, the algorithm of Prop. 2.3 is subexponential only if $\theta^* < 1$, which unfortunately seems beyond the current technology to prove, even under GRH. We can, however, make a reasonable conjecture of the value of θ^* by

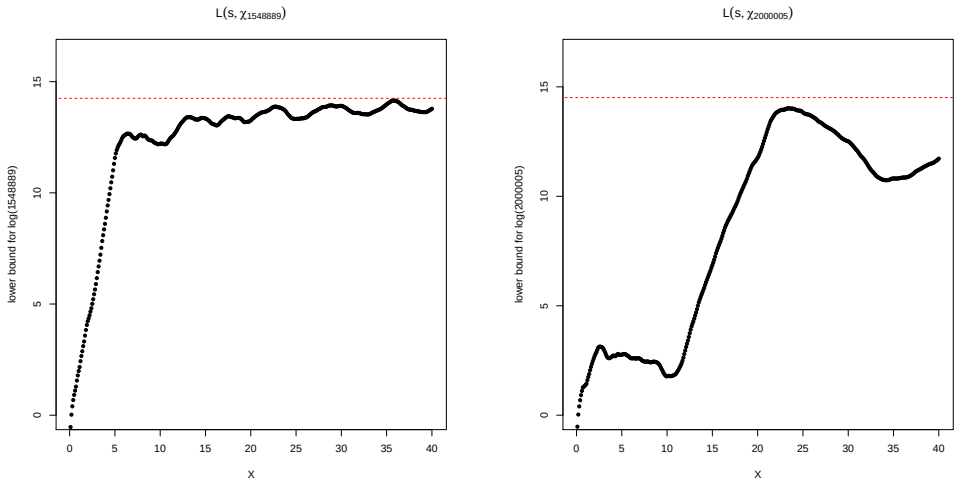

FIGURE 2. Behavior of the lower bound (4) as X increases: The case of a large gap (left) compared with the case of an average gap.

answering the analogous question for a suitable random matrix model, where the calculation is more tractable:

Conjecture 3.1. *We have*

$$\eta_\infty(\theta) = \begin{cases} 1 - \frac{\theta}{2} & \text{if } 0 < \theta < 1, \\ \frac{1}{2} & \text{if } \theta \geq 1. \end{cases}$$

In particular, $\theta^* = \frac{2}{3}$.

Thus, by Prop. 2.3, we conjecture that our algorithm is capable of certifying N squarefree in time $O\big(\exp[(\log N)^{2/3+o(1)}]\big)$.

We give a detailed justification for the conjecture in §3.1 below. First, however, it turns out that one can arrive at the same conclusion for the running time without any consideration of the zero sum in (3), by analyzing the lower bound (4) using a simple model of the $\chi_{q\Delta}(p)$ as independent random variables assuming the values 1 and -1 with equal probability (this is not always a good model [27] but suffices for our purposes). We make this more precise in the following proposition, which is a consequence of [21, Theorem 1].

Proposition 3.2. *Let* $Y_1, Y_2, \ldots$ *be independent random variables such that* $\mathbb{P}(Y_j = 1) = \mathbb{P}(Y_j = -1) = \frac{1}{2}$, *and put* $Y := 2\sum_{p_j \leq e^X} \frac{Y_j \log p_j}{\sqrt{p_j}} \Big(1 - \frac{\log p_j}{X}\Big)$, *where* p_j *denotes the* j*th prime number. Then, for each* n *satisfying* $3 \leq n < e^X$, *we have*

$$\mathbb{P}(Y \geq v_n) \geq 2^{-22} \exp\left(-\frac{30 v_n^2}{c_n}\right), \qquad \mathbb{P}(Y \geq u_n) \leq \exp\left(-\frac{u_n^2}{32 c_n}\right),$$

where $v_n := \sum_{p_j \leq n} \frac{\log p_j}{\sqrt{p_j}} \Big(1 - \frac{\log p_j}{X}\Big)$, $u_n := 4v_n$, *and* $c_n := \sum_{n < p_j \leq e^X} \frac{\log^2 p_j}{p_j} \Big(1 - \frac{\log p_j}{X}\Big)^2$.

In particular, as $n, X \to \infty$ with $n = e^{o(X)}$, so that $v_n \sim 2\sqrt{n}$ and $c_n \sim \frac{1}{12}X^2$, we get $\mathbb{P}(Y \geq 2\sqrt{n}) \geq \exp(-(1440 + o(1))n/X^2)$. Therefore, after $\lfloor \exp X \rfloor$ independent samples of Y, we expect to occasion $Y \gtrsim \frac{1}{6\sqrt{10}} X^{3/2}$ at least once. In the opposite direction, we

have $\mathbb{P}(Y \geq 8\sqrt{n}) \leq \exp(-(24+o(1))n/X^2)$, and so after $\lfloor \exp X \rfloor$ independent samples, we expect at most one instance of $Y \gtrsim \frac{4}{\sqrt{6}} X^{3/2}$. Together, these estimates are consistent with $\theta^* = 2/3$. Of course, Prop. 3.2 simplifies the situation by ignoring the higher prime powers, but that is not important since they contribute only $O(X)$, and so do not impact the $X^{3/2}$ term. It is worth noting, however, that the contribution of the higher prime powers in numerical computations is still noticeable because $\chi_q(p^2) = 1$ whenever $(q,p) = 1$, and so the bulk of their contribution is guaranteed to help our lower bound, regardless of the number of samples.

3.1. **A conjecture for θ^* via random matrix theory.** The random matrix philosophy suggests (e.g. by comparing the 1-level densities) that the relevant symmetry for a family of primitive quadratic twists is symplectic. The symplectic group $USp(2N)$ is a compact group consisting of $2N \times 2N$ unitary matrices A satisfying $A^t J A = J$, where

$$J := \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix}.$$

The eigenvalues of A lie on the unit circle, come in conjugate pairs, and can be written uniquely as

$$e^{\pm i\theta_1(A)}, \dots, e^{\pm i\theta_N(A)}, \qquad 0 \leq \theta_1(A) \leq \cdots \leq \theta_N(A) \leq \pi.$$

Making the identification $2N = \log|\Delta|$,[10] we expect that statistics of the lowest eigenphase $\theta_1(A)$ as A varies in $USp(2N)$ coincide to leading order, and modulo arithmetic effects, with statistics of the lowest zero $\gamma_1(q\Delta)$ as q varies in $\mathcal{F}$ but still sufficiently small compared to $|\Delta|$. Thus, by computing statistics of $\theta_1(A)$, we arrive at conjectures for $\gamma_1(q\Delta)$. In particular, since the complexity of our algorithm depends on the frequency of large values of $\gamma_1(q\Delta)$, we are led to consider the tail distribution of $\theta_1(A)$.

To this end, and to facilitate comparison with other symmetry groups later on, let $U(N)$ denote the (compact) group of $N \times N$ unitary matrices, and $SO(2N) \subset U(2N)$ the group of orthogonal matrices of determinant 1. The eigenphases of $A \in U(N)$ can be written uniquely as $0 \leq \theta_1(A) \leq \cdots \leq \theta_N(A) < 2\pi$, while those of $A \in SO(2N)$, which come in pairs $\pm\theta_j(A)$, can be written uniquely as $0 \leq \theta_1(A) \leq \ldots \leq \theta_N(A) \leq \pi$. Let $\mathbb{P}_{G(N)}$ denote the unique Haar measure on $G(N) \in \{U(N), SO(2N), USp(2N)\}$, normalized to be a probability measure. The random matrix philosophy suggests, for example, that the relevant symmetry group for averages over a family of twists by n^{it} is unitary, while for averages over a family of elliptic curves it is orthogonal (even or odd, depending on the sign of the functional equation in the family).

Let $\mathbb{P}_{G(N)}^{\times M} = \mathbb{P}_{G(N)} \times \cdots \times \mathbb{P}_{G(N)}$, repeated M times, be the product measure on $G(N)^M$. For each Borel-measurable set $J \subset [0, \sigma\pi]$, where $\sigma = 2$ if $G(N) = U(N)$ and $\sigma = 1$ otherwise, define $\mathcal{S}(J) := \{(A_1, \dots, A_M) \in G(N)^M : \max_{1\leq m\leq M} \theta_1(A_m) \in J\}$. For short-hand, we write $\mathbb{P}_{G(N)}(\max_{1\leq m\leq M} \theta_1(m) \in J)$ in place of $\mathbb{P}_{G(N)}^{\times M}(\mathcal{S}(J))$, $\mathbb{P}_{G(N)}(\max_{1\leq m\leq M} \theta_1(m) > s)$ in place of $\mathbb{P}_{G(N)}^{\times M}(\mathcal{S}((s,\infty)))$, and so on. The distribution function $\mathbb{P}_{G(N)}\big(\theta_1 > s\big)$ is known as the gap probability. The proofs of Propositions 3.3–3.4 below are given in the appendix.

[10] This identification is obtained by equating the mean spacing of eigenphases of $A \in USp(2N)$, which is π/N, and the mean spacing of zeros of $L(s, \chi_{q\Delta})$ at a fixed height, which is $\sim 2\pi/\log|q\Delta| \sim 2\pi/\log|\Delta|$ as $|\Delta| \to \infty$, $\Delta \in \mathcal{F}$.

Proposition 3.3. *Fix $\beta \in (0,2)$, and define*

$$M_\beta(N) := \left\lfloor \exp((2N)^\beta) \right\rfloor, \quad s^\pm_{\varepsilon,\beta}(N) := (4 \pm \varepsilon)(2N)^{\beta/2-1}.$$

Then, for each fixed $\varepsilon > 0$, as $N \to \infty$ we have

$$\mathbb{P}_{USp(2N)}\left(s^-_{\varepsilon,\beta}(N) < \max_{1 \le m \le M_\beta(N)} \theta_1(m) \le s^+_{\varepsilon,\beta}(N) \right) \to 1.$$

In other words, $(2N)^{1-\beta/2} \max_{1 \le m \le M_\beta(N)} \theta_1(m)$ converges in distribution to 4.

Therefore, if we choose $A_1, \dots, A_{M(N)} \in USp(2N)$, independently and uniformly with respect $\mathbb{P}_{USp(2N)}$, then in the limit as $N \to \infty$, we have $\max_{1 \le m \le M(N)} \theta_1(A_m) > s^-_{\varepsilon,\beta}(N)$ with probability approaching 1. For instance, if $\beta = 1$, then we expect to find at least one lowest eigenphase of size $\gtrsim (4-\varepsilon)/\sqrt{2N}$. Since the eigenvalues of symplectic matrices come in conjugate pairs, this corresponds to an eigenphase spacing $\ge 2(4-\varepsilon)/\sqrt{2N} \approx \sqrt{32/N}$, which is $\frac{4}{\pi}\sqrt{2N}$ times the average spacing.

In contrast, the eigenvalues of unitary matrices do not necessarily come in conjugate pairs, so the point 1 on the unit circle is no longer distinguished, and actually $\mathbb{P}_{U(N)}$ is rotationally invariant. Thus, it is more natural to consider the nearest-neighbor distribution function, $\mathbb{P}_{U(N)}(\theta_2 - \theta_1 > u) := \mathbb{P}_{U(N)}(\{A \in U(N) : \theta_2(A) - \theta_1(A) > u\})$, which is related to the gap probability by differentiation; see (23) in the appendix. In fact, $\log \mathbb{P}_{U(N)}(\theta_2 - \theta_1 > u) \sim \log \mathbb{P}_{U(N)}(\theta_1 > u)$ as $N \to \infty$ over a wide range of u. To facilitate comparison with the symplectic case, we consider the half-spacings $\frac{1}{2}[\theta_2 - \theta_1]$ in the following.

Proposition 3.4. *Fix $\beta \in (0,2)$, and define*

$$M_\beta(N) := \left\lfloor \exp(N^\beta) \right\rfloor, \quad u^\pm_{\varepsilon,\beta}(N) := \sqrt{8 \pm \varepsilon} N^{\beta/2-1}.$$

Then, for each fixed $\varepsilon \in (0,8]$, as $N \to \infty$ we have

$$\mathbb{P}_{U(N)}\left(u^-_{\varepsilon,\beta}(N) < \max_{1 \le m \le M_\beta(N)} \tfrac{1}{2}[\theta_2(m) - \theta_1(m)] \le u^+_{\varepsilon,\beta}(N) \right) \to 1.$$

In other words, $N^{1-\beta/2} \max_{1 \le m \le M_\beta(N)} \frac{1}{2}[\theta_2(m) - \theta_1(m)]$ converges in distribution to $\sqrt{8}$. The same is true if, in addition, one maximizes over the spacings of each matrix, replacing $\frac{1}{2}[\theta_2(m) - \theta_1(m)]$ by $\max_{1 \le j \le N} \frac{1}{2}[\theta_{j+1}(m) - \theta_j(m)]$, where $\theta_{N+1} := 2\pi + \theta_1$.

In light of Prop. 3.4, we see that sampling from $U(2N)$ does not do as well as sampling from $USp(2N)$. For if we choose $\lfloor \exp((2N)^\beta) \rfloor$ matrices from $U(2N)$, independently and uniformly with respect to $\mathbb{P}_{U(2N)}$, then we expect that half the max spacing is $\approx \sqrt{8}(2N)^{\beta/2-1}$, which is worse than the symplectic case by a factor of $\sqrt{2}$. Note that we could have compared $USp(2N)$ and $U(N)$ instead, but to do so meaningfully the eigenphases in the two ensembles should be re-normalized to have the same mean spacing, so that $U(N)$ still does worse by a factor of $\sqrt{2}$.

This suggests that our algorithm should do better if it searches through quadratic twists rather than twists by n^{it}, i.e. $\gamma_1(q\Delta)$ as opposed to $\frac{1}{2}[\gamma_{j+1}(\Delta) - \gamma_j(\Delta)]$, for q and j in a suitable range.[11] This agrees with our observations in practice. An additional reason for it might be that the assumption of independent samples is less applicable to the gaps $\gamma_{j+1}(\Delta) - \gamma_j(\Delta)$,

[11] To clarify the analogy with $U(N)$ a little more, we expect $\max_{t \le \gamma_j(\Delta) < t+2\pi} \frac{1}{2}[\gamma_{j+1}(\Delta) - \gamma_j(\Delta)]$, for $t = |\Delta|^{o(1)}$, to be modelled by $\max_{1 \le j \le N} \frac{1}{2}[\theta_{j+1}(m) - \theta_j(m)]$, where $N \approx \log|\Delta|$.

since they come from a single L-function, and are thus constrained by its analytic properties, in contrast to the $\gamma_1(q\Delta)$, which come from different L-functions. For example, we intuitively expect $\gamma_{j+1}(\Delta) - \gamma_j(\Delta)$ to have negative correlations over short ranges (and also some long-range correlations due to the primes; see [22] for a numerical discussion of this in the case of zeta). While such negative correlations do not affect the 2/3 in the analogue of Conj. 3.1 for the t-aspect, they likely make the implied asymptotic constants worse.

In order to make a conjecture for θ^* based on our $USp(2N)$ calculation, we identify $2N$ with $\log|q\Delta|$, as usual, and the lowest eigenphase with $\gamma_1(q\Delta)$. If $\theta < 1$ then twisting by χ_q does not affect the density of zeros appreciably, so we may interpret Prop. 3.3 for fixed $2N \approx \log|\Delta|$ as sampling $\gamma_1(q\Delta)$ for q from $\{q \in \mathcal{F} : (q,\Delta) = 1, |q| \le \exp((\log|\Delta|)^\theta)\}$. The conclusion of the proposition thus suggests that $M_\Delta(\theta) \asymp (\log|\Delta|)^{\theta/2-1}$; in particular, we expect $\eta_\infty(\theta) = 1 - \theta/2$, and so $\theta^* = 2/3$. On the other hand, if $\theta > 1$ then q becomes the dominating factor in the zero density; thus, we expect the maximum of $\gamma_1(q\Delta)$ to be attained for a relatively small choice of q, meaning we do not derive any benefit from increasing θ further, and $\eta_\infty(\theta)$ is constant. Note that similar conclusions are reached if we sample twists by n^{it} instead.

Finally, we remark that Conj. 3.1 is of independent interest and may warrant further study. One can try to confirm it directly (i.e. by computing the first zero of many twists), but this requires taking $|\Delta|$ fairly large before one can hope to discern a clear pattern. Basic experiments suggest taking $|\Delta| \gtrsim 10^{15}$, say, which is prohibitively time-consuming using the standard approximate functional equation, as one would need to compute $\gamma_1(q\Delta)$ for millions of q. It would perhaps be better to formulate a precise conjecture for $M_\Delta(\theta)$ itself, including lower order terms, and check the numerics for that. One could also try to confirm the conjecture for other families.

3.2. **Numerical results.** We applied our method to several `RSA`-numbers, but our main test case was `RSA-210`, which is the following 210-digit number:

$$\begin{aligned}\texttt{RSA-210} = {} & 24524664490027821197651766357308801846702678767833327\\ & 59743414451715061600830038587216952208399332071549 10\\ & 36268271916798640797767232430056005920356312465612 18\\ & 46581790410013185929961993381701214933503487587055 1067.\end{aligned}$$

We searched for candidate twists (i.e. twists that are expected to make the prime sum large) essentially by brute force, with some modest refinements described in §4.2.1. We first used a simple weighting function, such as a triangle wave, to evaluate a short prime sum (typically with $p \le 10^4$) for all twists within a given range, then incrementally increased the length of the sum as we filtered the results. The candidates found this way were then fed into the lower bound (4), this time using a much longer prime sum and the test function produced by the quadratic form method outlined in §2.1.[12] Our best-performing twist was

[12]Note that the integral in (3) can be computed to high precision using standard numerical integration methods. Moreover, in our final, long prime sum, we used approximations of the test function by Chebyshev polynomials, which allows most of the summation to be carried out in integer arithmetic. In this way we can effectively control the round-off error in the computation. On the other hand, since the longest sum that we computed was over the primes $\le 2.5 \times 10^{16}$, standard double-precision arithmetic would also suffice to control the round-off errors effectively for many choices of g, e.g. by using pairwise summation.

$-9334602088654580277283 = -568391 \times 2345033 \times 7003250461$, which yielded the lower bound $\log|\Delta| \geq 137.5158$ using $X = \log(1.3 \times 10^{15})$ and $M = 312$ in (5).

Proof of Theorem 1.1. We illustrate the proof for the case of $N =$ `RSA-210`; the parameter choices that we used to complete the proof for the other challenge numbers are summarized in Table 1.

Suppose, to the contrary, that all prime factors of N have multiplicity > 1. Then $N = s^2|\Delta|^3$, where $|\Delta|$ is the conductor of χ_{-N}. We verified that N is not a perfect cube, and our computation showed that $\log|\Delta| > 137.515$, so that $1 < s < 1.3 \times 10^{15}$. This leads to a contradiction since, as a by-product of computing the prime sum in the explicit formula, we checked that N has no non-trivial factor $< 1.3 \times 10^{15}$. □

N	q	e^X	M	$\log\|\Delta\| \geq$
`RSA-210`	$-93346020886545802772 83$	1.3×10^{15}	312	137.5158
`RSA-220`	9700641183360814771 09	1.8×10^{15}	312	145.2599
`RSA-230`	22981707927294468438 01	2.5×10^{16}	312	150.8289
`RSA-232`	$-27792634603676954310 79$	10^{16}	312	152.7847

TABLE 1. Parameters used in the proof of Theorem 1.1

Remark. We do not need the full strength of the bound $\log|\Delta| > 137.515$ to prove the theorem, as we separately ruled out factors $\leq 10^{20}$ using the implementation of Pollard's $p-1$ method[13] in GMP-ECM [10], which takes less than a day on a computer with 80GB of memory. Therefore, the bound $\log|\Delta| > 130.02$ suffices to prove the theorem, which reduces the size of the prime sum needed to $p \leq 2.66 \times 10^{14}$. A similar improvement was noted for the other challenge numbers.

In Figure 3, we present data about the practical efficiency of our algorithm, providing further evidence for the 2/3 exponent. To clarify the situation, recall that θ^* is chosen to balance the number of terms in the prime sum versus the number of twists we need to try so that, with high probability, the zero contribution is small for at least one twist. We accomplish this by taking g with support inversely proportional to the largest gap that we anticipate after trying $\exp((\log|\Delta|)^\theta)$ twists, and so our prime sum has length $\leq \exp(c_1(\log|\Delta|)^{\eta_\Delta(\theta)})$ for some suitable constant $c_1 > 0$. The 2/3 arises as Prop. 3.3 suggests that $\eta_\Delta(\theta) \approx 1 - \theta/2$, and on solving $1 - \theta/2 = \theta$. More precisely, it arises since if we sample $\gamma_1(q\Delta)$ over $q \in \mathcal{F}$, $|q| < \exp X$, with X much smaller than $\log|\Delta|$, then Prop. 3.3 suggests we should encounter at least one $\gamma_1(q\Delta) \geq 4\sqrt{X}/\log|\Delta|$. Therefore, looking back at Prop. 2.2, we expect to obtain a lower bound very close to $\log|\Delta|$ in time $\lesssim \exp(X + \frac{c_2 \log|\Delta| \log\log|\Delta|}{4\sqrt{X}})$, where $c_2 > 0$ is a constant implied by the proposition. Optimizing, we choose $X = (\frac{c_2}{4}\log|\Delta|\log\log|\Delta|)^{2/3}$.

This reasoning on its own does not fully explain what we observe in Figure 3, which is that by sampling $\exp X$ twists and using a prime sum of length $\exp X$, we seem to obtain a lower bound like $X^{3/2}$, even for intermediate values of X much smaller than $(\log|\Delta|)^{2/3}$. This

[13]Like Pollard–Strassen, this method can be used to rule out small factors, with comparable complexity. Pollard–Strassen has a theoretical advantage, in that the $p-1$ method produces an inconclusive result if a randomly-chosen residue happens to have exactly the same order modulo every prime factor of N; however, the chance of that occurring is vanishingly small, so this is irrelevant in practice.

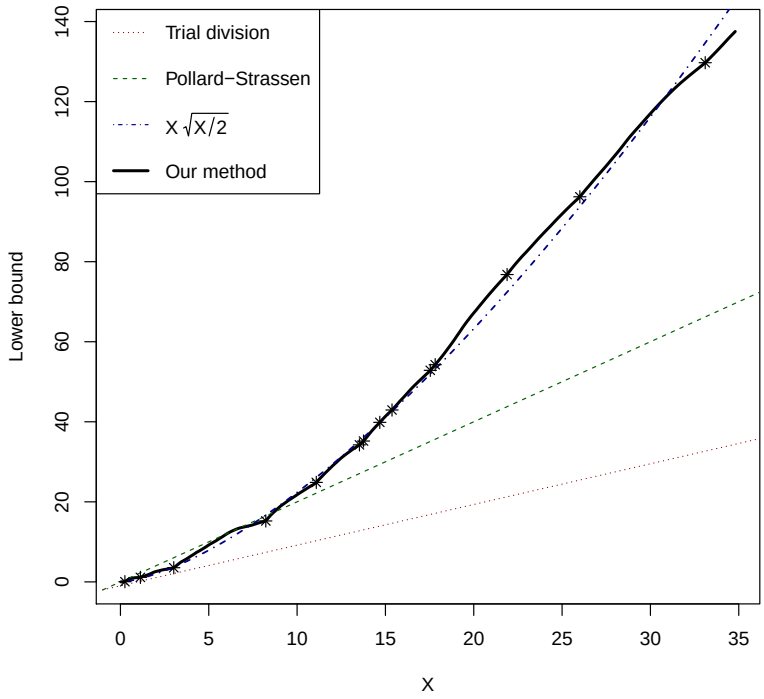

FIGURE 3. The lower bound (under GRH) produced by our method when applied to `RSA-210` using the primes $\le e^X$. The $*$s mark the places where the best performing twist available so far changes. The slope increases noticeably at each $*$, except towards the end, where it is likely that we are not finding the best twists. Also note that the $\sqrt{2}$ in our fitted curve is likely not an absolute constant, but varies as a small power of $\log\log|\Delta|$.

behavior is expected if one treats the prime sum as a sum of independent random variables, as in Prop. 3.2, but it would be reassuring to see it from the zeros directly. The difficulty towards this is that if h does not have sufficient decay outside the large gap, then we cannot bound the contribution of the zeros effectively (cf. Prop. 2.2). Nevertheless, one can obtain a heuristic explanation, as follows.

We choose h with $0 \le h(t) \le 1$, say, and mostly concentrated within the interval $[-1/X, 1/X]$, roughly speaking. We let $N_\chi(t) := \#\{0 \le \gamma(\chi) < t\}$ denote the zero-counting function, and assume $L(1/2,\chi) \ne 0$ for simplicity. Then $\sum_{\gamma(\chi)} h(\gamma(\chi)) = 2\int_0^\infty h(t)\,dN_\chi(t)$. The contribution of the smooth part of $N_\chi(t)$ to the integral is $\sim g(0)\log|\Delta|$, which is precisely the left-hand side of (3). Therefore, the prime sum contribution, which is basically our lower bound, should be $\approx -2\int_0^\infty h(t)\,dS_\chi(t)$, where $S_\chi(t)$ is the fluctuating part of $N_\chi(t)$. This last integral is typically very small due to the random nature of S_χ, except we purposefully introduced a bias in it via our choice of twist, resulting in a large gap around the center of size like $\sqrt{X}/\log|\Delta|$. The contribution of this bias to the prime sum is essentially, for a reasonable h, $-2\int_0^{\sqrt{X}/\log|\Delta|} h(t)\,dS_\chi(t) \gg \sqrt{X}$. Since we expect the contribution from the interval $[\sqrt{X}/\log|\Delta|, \infty)$ to wash out in comparison,[14] we should get a lower bound like $g(0)\log|\Delta| \gg \sqrt{X}$. Finally, since $g(0) \ll 1/X$, we should get $\log|\Delta| \gg X^{3/2}$.

[14]This is the part of the heuristic that we cannot prove, even under the GRH, unless h has sufficient decay outside the zero gap.

This heuristic indicates how the running time of our algorithm is controlled by extreme (negative) values of $S_\chi(t)$. If $S_\chi(t) \ll_t (\log|\Delta|)^{1/2+o(1)}$, for example, then we cannot expect a running time better than $\exp((\log|\Delta|)^{1/2+o(1)})$, even if we allow for an oracle supplying the algorithm with the best twist in any requested range. (This is in agreement with Conj. 3.1.) On the other hand, if $S_\chi(t)$ can get much larger (without violating the GRH, so our method can still apply!), then there is no such barrier.

4. Refinements

In this section, we describe a few refinements of our basic method and indicate some directions for future research.

4.1. **Linear programming.** A natural question is whether one can make better use of the zero sum in (3) than simply ignoring it by positivity, as in (4), especially since it typically dominates the right-hand side when X is small. One idea is to apply the explicit formula (3) with various choices of test function, setting up a system of inequalities, and try to obtain a non-trivial lower bound for the sum over zeros. Since $\log|\Delta|$ also appears in (3) and remains unknown to us, the logic of this may seem circular at first glance, but we gain some additional information coming from the fact that the zeros occur discretely, as we elaborate below.

An immediate practical problem is that the system involves infinitely many variables, since the zero sum is infinite, and h cannot be compactly supported (it has to be analytic). Nevertheless, one can reduce to a finite number of variables, without too much loss, using an explicit estimate of the form $|\sum_{|\gamma|\geq T} h(\gamma)| = |2\int_T^\infty h(t)\,dN_\chi(t)| \leq \mathcal{E}(h,T)$, $T>0$, simply bounding the conductor by the modulus, and using known estimates for $S_\chi(t)$. Hence

$$\sum_{|\gamma|<T} h(\gamma) - \mathcal{E}(h,T) \leq \sum_{\gamma} h(\gamma) \leq \sum_{|\gamma|<T} h(\gamma) + \mathcal{E}(h,T). \tag{8}$$

A more serious problem is that the system is not linear in the zero ordinates, and therefore is likely very unstable. We linearize the system, at the cost of having more variables or extra solutions, by subdividing the interval $[0,T)$ into bins of size δ, so that the variables become the count of zeros in each bin rather than the zeros themselves. Specifically, for each integer $V>0$, and each integer $v\in[0,V)$, let $\delta := T/V$, $I(v) := [v\delta,(v+1)\delta)$, $m(v) := \frac{1}{2}\#\{\gamma : |\gamma|\in I(v)\}$, $h^+(v) := \sup_{t\in I(v)} h(t)$, and $h^-(v) := \inf_{t\in I(v)} h(t)$. Then we have

$$2\sum_{0\leq v<V} m(v)h^-(v) \leq \sum_{|\gamma|<T} h(\gamma) \leq 2\sum_{0\leq v<V} m(v)h^+(v)\,. \tag{9}$$

Applying (8) and (9) with a set of test functions $\{(g_k,h_k) : 1\leq k\leq K\}$ of our choice, we obtain a linear system

$$\begin{aligned} 2\sum_{0\leq v<V} m(v)h_k^-(v) - \mathcal{E}(h_k,T) &\leq g_k(0)\log|\Delta| + g_k(0)\log q - P(g_k,q) \\ &\leq 2\sum_{0\leq v<V} m(v)h_k^+(v) + \mathcal{E}(h_k,T) \end{aligned} \tag{10}$$

for $k=1,\dots,K$, where χ_q is the twist used, and $P(g_k,q)$ denotes the contribution from the prime sum and integral terms in (3). Note that V controls the size of each bin, and T controls the point where we truncate the zero sum. Finally, we let `logd` denote the unknown

value of $\log|\Delta|$, and feed the system (10) into a linear programming solver, such as `GLPK` [11], with `logd` as the objective function to be minimized.

We experimented with this approach for `RSA-210` using various choices of q, T, V, and $\{(g_k, h_k) : 1 \le k \le K\}$. For example, one of the better performing twists found, as in §3.2, was $q = -65123121667$. Using this twist, we set up the system (10) with $T = 4$ and $V = 500$ (so that $\delta = 0.008$, which is smaller than the mean zero spacing ≈ 0.013), and

$$h_k(t) = \left[\frac{\sin(Xt/2k)}{(Xt/2k)}\right]^{2k}, \qquad k = 1, \dots, 7, \qquad X = 7\log 10,$$

so that $g_k(x)$, $k = 1, \dots, 7$, are supported on $|x| \le X$.[15] We imposed an integer variable constraint on $m(v)$, $v = 0, \dots, 44$, with the rest being real variables. The integer variables are located at the beginning, covering the interval $[0, 0.36)$, which is reasonable since $h_k(t)$ is not too small there and so detected zeros have more weight. Solving this system, we obtained the lower bound $\log|\Delta| \ge 47.153$, of which 2.494 came from the zeros. This represents an improvement of about 5.5% over using $\max_{1\le k\le 7}[P(g_k, q) - g_k(0)\log q]$ alone, which is comparable to the improvement that we obtained from using the Pollard $p-1$ algorithm to rule out small values of ℓ, as remarked after the proof of Thm. 1.1. Although this is a modest improvement on a logarithmic scale, it makes a substantial difference in the length of the final prime sum.

In general, further gains are possible by using more integer variables, a smaller grid spacing (smaller δ), or additional test functions, in that order of importance. In reality, adding more test functions of compact support of size X loses impact quickly, which is not surprising because such functions cannot resolve zeros to better than $O(1/X)$. Most of the gains, in fact, come from imposing integer constraints. If no integer constraints are imposed, the improvement in the above example goes down significantly, to around 1%. Also, if all the variables are real, then the linear programming approach is closely related to the approach of varying the test function, described in §2.1, and so cannot be expected to do significantly better.[16]

However, one has to weigh the extra time it takes to set up and solve the mixed integer programming problem against the time it takes to simply compute a longer prime sum. In the above example, it took about 15 minutes to solve the problem, but it can take much longer if more integer constraints are imposed. It is tempting to think that if one could allow the number of integer variables to grow very large without significant time penalty then there would be no limit to the improvement that could be obtained. We offer the following theoretical evidence in favor of that belief.

Definition 4.1. *Let $S = \{z \in \mathbb{C} : |\Im(z)| < 1/2\}$. A* divisor *on S is a function $m : S \to \mathbb{Z}$ which is supported on a discrete subset of S. A divisor m is* admissible *if $m(-\gamma) = m(\gamma) \ge 0$ for all $\gamma \in S$ and there is a number $A \ge 0$ such that $\sum_{\substack{\gamma\in S\\ |\gamma|\le T}} m(\gamma) \ll T^A$ for all $T \ge 1$.*

[15] The inequalities in (10) were imposed in both directions except for h_1, where only the lower bound was used.

[16] In the real variable case one can obtain an easily verifiable certificate that the solution is indeed correct by solving the dual problem. This is not available if one imposes integer constraints, and so one has to trust the linear programming software in that case.

Proposition 4.2. *Let $m : S \to \mathbb{Z}_{\geq 0}$ be an admissible divisor, $d \in \mathbb{R}^\times$, and $\{c_n\}_{n=2}^\infty$ a sequence of complex numbers satisfying $c_n \ll n^{-\delta}$ for some $\delta > 0$. Suppose that for every smooth, even function $g : \mathbb{R} \to \mathbb{C}$ of compact support and cosine transform h we have the equality*

$$g(0)\log|d| = \sum_{\gamma\in S} m(\gamma)h(\gamma) + 2\sum_{n=2}^{\infty} c_n g(\log n) + g(0)\log(8\pi e^\gamma) - \int_0^\infty \frac{g(0)-g(x)}{2\sinh(x/2)}\,dx + (\operatorname{sgn} d)\int_0^\infty \frac{g(x)}{2\cosh(x/2)}\,dx.$$

Then d is a fundamental discriminant, $c_n = \frac{\Lambda(n)\chi_d(n)}{\sqrt{n}}$ for every $n \geq 2$, and $m(\gamma) = \operatorname{ord}_{s=1/2+i\gamma} L(s,\chi_d)$ for all $\gamma \in S$.

Thus, the explicit formula is rigid in the sense that the only identities of the shape (3) that can hold for all test functions are the ones arising from quadratic character L-functions. We remark that the key to this proposition, whose full proof is given in the appendix, is that m is integer valued and supported on a discrete set. Unfortunately, the proposition is ineffective, in that it does not predict how many or how complicated we must choose the test functions before finding a system that yields a good lower bound for $\log|d|$. However, note that under GRH, the $\Delta \in \mathcal{F}$ with $|\Delta| \leq x$ are distinguished from one another by the values of $\chi_\Delta(p)$ at primes $p \leq O(\log^2 x)$ [17]. This statement alone does not offer any indication of how to find Δ given a list of its initial character values, but together with Prop. 4.2 it suggests that a given Δ might be captured by the system (10) using test functions supported up to $X \approx 2\log\log|\Delta|$, provided that we are allowed to take V and K arbitrarily large. However, our numerical experiments so far, which were limited to at most a few hundred integer variables, have not corroborated this speculation, even allowing for larger values of X.

4.2. **Finding correlating characters.**

4.2.1. *Lining up the initial primes.* In order to improve the efficiency of the brute force search, we chose q so as to line up the values of the prime sum for small n, i.e. so that $\chi_{q\Delta}(p) = 1$ for small primes p. Of course there is no guarantee that doing so is optimal, and indeed it is likely that the best choices of q of a given size adhere to this principle only loosely, i.e. they may sacrifice a few small values of p in order to line up many more. However, if we have the resources to evaluate the prime sum for a fixed number of q, regardless of size (a reasonable assumption, since the only operation performed on q itself is reduction mod p), then it makes sense to line up the small primes in attempt to skew the distribution of values in our favor.

To be more precise, consider an idealized form of the lower bound (6) with h a δ-function and $g \equiv 1$. Then for a prime power $n = p^k$, the corresponding term of (6) is $2\chi_{q\Delta}(n)\Lambda(n)/\sqrt{n}$. The expected value of this term, that is its average value over all $q \in \mathcal{F}$, is easily seen to be 0 if k is odd and $2\frac{p}{p+1}\frac{\Lambda(n)}{\sqrt{n}}$ if k is even.[17] Thus, if we force q to satisfy $\chi_{q\Delta}(p) = 1$, this

[17] $\mathbb{E}(\chi_{q\Delta}(p^{2k})) = \mathbb{E}(\chi_{q\Delta}(p^2)) = \frac{\phi(p^2)}{p^2-1} = \frac{p}{p+1}$, where the second equality holds because $q \in \mathcal{F}$, so that $q \not\equiv 0 \pmod{p^2}$.

introduces a positive bias in the prime sum of

$$\sum_{k=1}^{\infty} 2\frac{\Lambda(p^k)}{p^{k/2}} \begin{cases} \frac{1}{p+1} & \text{if } 2 \mid k, \\ 1 & \text{if } 2 \nmid k \end{cases} = \frac{2\log p}{p-1}\left(\sqrt{p} + \frac{1}{p+1}\right).$$

However, it also comes with a price, in that we expect such a q to be about $2(p+1)/p$ times larger than a fundamental discriminant chosen randomly without regard to the value of $\chi_q(p)$. Thus, our expected net improvement is

$$\frac{2\log p}{p-1}\left(\sqrt{p} + \frac{1}{p+1}\right) - \log\frac{2(p+1)}{p}. \tag{11}$$

(A similar argument applies to forcing $\chi_{q\Delta}(-1) = 1$, from which we expect a net improvement of $\frac{\pi}{2} - \log 2$.) It turns out that (11) is positive for $p \le 251$ but negative for larger primes.

4.2.2. *The shortest lattice vector problem.* It is plausible that there is a better strategy for finding good twists than a brute-force search, meaning a strategy that can find the same quality twist as brute force but using much less sampling. If one could be assured of finding $\gamma_1(q\Delta) \gg \sqrt{X}/\log|\Delta|$ in a subset of $\{q \in \mathcal{F} : (q,\Delta) = 1, |q| \le \exp X\}$ of size $\ll \exp(X^\tau)$, $0 \le \tau < 1$, then one could improve the 2/3 exponent to $\max\{\theta^*_{\min}, \frac{2\tau}{2\tau+1}\}$, where $\theta^*_{\min} := \inf_{\theta>0}\eta_\infty(\theta)$, provided the subset can be determined easily. An obvious candidate is the subset of smooth fundamental discriminants. For example, one could search for a product of real primitive characters $\chi_q = \chi_{q_1}\cdots\chi_{q_m}$, $|q_j| < Q$, $q_j \ne q_k$, that correlates strongly with χ_d, so as to make

$$2\sum_{p<P}\frac{\log p}{\sqrt{p}} - 2\sum_{p<P}\frac{\chi_q(p)\chi_d(p)\log p}{\sqrt{p}} + \log|q| \tag{12}$$

small, in the hope that it will lead to an unusually large prime sum in the explicit formula. The question of finding a good choice of χ_q can be framed in terms of finding a short vector in the lattice generated by the rows of the following $(n+m+1)\times(n+m+1)$ matrix, as we explain next. (This idea was applied in [23] in the t-aspect to disprove the Mertens conjecture.)

$$\begin{bmatrix}
i(d,p_1) & i(d,p_2) & \cdots & i(d,p_n) & 0 & 0 & \cdots & 0 & 2^M \\
i(q_1,p_1) & i(q_1,p_2) & \cdots & i(q_1,p_n) & \lfloor 2^M\sqrt{\log|q_1|}\rfloor & 0 & \cdots & 0 & 0 \\
i(q_2,p_1) & i(q_2,p_2) & \cdots & i(q_2,p_n) & 0 & \lfloor 2^M\sqrt{\log|q_2|}\rfloor & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
i(q_m,p_1) & i(q_m,p_2) & \cdots & i(q_m,p_n) & 0 & 0 & \cdots & \lfloor 2^M\sqrt{\log|q_m|}\rfloor & 0 \\
2w(p_1) & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 2w(p_2) & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 2w(p_n) & 0 & 0 & \cdots & 0 & 0
\end{bmatrix}$$

$$w(p) := \left\lfloor \frac{2^{M+1}\sqrt{\log p}}{p^{1/4}} \right\rfloor, \quad i(q,p) := \frac{1}{2}\left(1 + \chi_{q^*}(p)\right)w(p), \quad q^* = \begin{cases} (-1)^{\frac{q-1}{2}}q & \text{if } q \text{ odd prime}, \\ q & \text{if } q \in \{-4, 8, -8\}. \end{cases}$$

Here, M is a large integer of our choice (in our application it was a random integer in $[75, 150)$). The weight $w(p)$ comes from (12), and indicates that it is more important to

correlate smaller primes. The weight $\sqrt{\log|q_j|}$ in the $(n+1)$st to $(n+m)$th columns indicates that using $\chi_{q_j^*}$ will incur a penalty imposed according to the explicit formula. The bottom n rows indicate that the kth entry in each row, $1 \le k \le n$, should be treated modulo $2w(p_k)$. Here, it is helpful to note that $i(q_j, p_k)$ is either 0 or 1 times $w(p_k)$, and so working modulo $2w(p_k)$ essentially means that only one multiple of each row is needed. Therefore, a vector in the lattice with a non-zero $(n+m+1)$-st entry, can be written in the form

$$\left(y_1 w(p_1), \dots, y_n w(p_n), u_1\sqrt{\log|q_1|}, \dots, u_m\sqrt{\log|q_m|}, 2^M\right), \tag{13}$$

where $y_k, u_j \in \{0,1\}$. The character generated by this vector is $\chi_{q_{\mathcal{J}}} := \prod_{j\in\mathcal{J}} \chi_{q_j^*}$, where $\mathcal{J} := \{1 \le j \le m, u_j = 1\}$, and it has discriminant $q_{\mathcal{J}} = \prod_{j\in\mathcal{J}} q_j^*$. The y_k are 0 or 1 according to whether $\chi_{\mathcal{J}}(p_k) = \chi_d(p_k)$ or not. Hence, in order for the vector (13) to be short, it means that

$$\sum_{\chi_{q_{\mathcal{J}}}(p_k)\neq\chi_d(p_k)} w(p_k)^2 + \sum_{j\in\mathcal{J}} \left\lfloor 2^M\sqrt{\log|q_j|}\right\rfloor^2 + 2^{2M} \approx 2^{2M}\left[4\sum_{\substack{\chi_{q_{\mathcal{J}}}(p_k)\neq\\ \chi_d(p_k)}} \frac{\log p_k}{\sqrt{p_k}} + \sum_{j\in\mathcal{J}} \log|q_j| + 1\right]$$

has to be small. This expression is essentially the same as (12), which we wish to minimize, but with $\chi_q = \chi_{q_{\mathcal{J}}}$ and $q = q_{\mathcal{J}}$. Thus, it is seen that one can find a good choice of χ_q if one can find a short vector in the lattice.

Finding the shortest vector in a lattice is conjectured to be NP-hard in the l^2-norm and the corresponding decision problem is conjectured to be NP-complete (see [1]). However, one can find relatively short vectors in polynomial time using the LLL algorithm of Lenstra, Lenstra, and Lovás [18], which produces a basis that is nearly orthogonal. The LLL algorithm was first used to factor a primitive univariate polynomial in polynomial time. It does not necessarily find the shortest vector, and it usually does not, but it can find relatively short vectors quickly.

We applied LLL to our lattice with P and Q ranging between 100 to over 1000. While it did yield above-average choices of χ_q, such as $q = -73147$, our best-performing twists ultimately came from the brute-force approach described in §4.2.1.

4.3. **More general twists.** If π is a cuspidal automorphic representation of $GL_r(\mathbb{A}_{\mathbb{Q}})$ with conductor q relatively prime to Δ, then the twist $\pi \otimes \chi_\Delta$ has conductor $q|\Delta|^r$. Assuming GRH for the associated L-function $L(s, \pi\otimes\chi_\Delta)$, we get a lower bound for $\log|\Delta|$ via the explicit formula. Thus, the idea of using twists as in §2.2 admits a vast generalization.

For any natural family of twists, one can expect a more general version of Prop. 2.1 to hold, i.e. for each $X > 0$ there will be some optimal choice of input data (π, g), where π is an element of the family and $g \in \mathcal{C}(X)$ is a test function to use in the explicit formula. For instance, considering the family of quadratic twists as in §2.2, it is easy to see that the right-hand side of (6) is bounded above by $O_X(1) - \log|q|$, uniformly for $g \in \mathcal{C}(X)$. Thus, only finitely many q are relevant, so it follows from Prop. 2.1 that there is a pair $(q, g) \in \mathcal{F}\times\mathcal{C}(X)$ which maximizes the lower bound (6).

Similarly, one can expect an analogue of Prop. 2.3 to hold for any given family. It would be of interest to study other families to see which yield the best performance. We conclude by listing a few families of L-functions that would make good candidates for future investigations.

- *Elliptic curve L-functions.* Can one make use of the BSD conjecture and the existence of high-order zeros at the central point to force zero repulsion?
- *Dedekind ζ-functions.* Can one make use of the existence of towers of number fields of bounded root discriminant?
- *Rankin–Selberg products.* Can one make use of the algebraic structure of the coefficients of L-functions to find correlating twists quickly?

Appendix A. Proofs

A.1. **Proof of Prop. 2.1.** Let $g_n \in \mathcal{C}(X)$, $n = 1, 2, 3, \ldots$, be a maximizing sequence for $l(g)$, with corresponding cosine transforms h_n. Since each h_n is non-negative, we have $|g_n(x)| \leq g_n(0) = 1$. Therefore, for $j \in \mathbb{Z}_{\geq 0}$,

$$\left|h_n^{(2j)}(0)\right| = \left|2(-1)^j \int_0^X x^{2j} g_n(x)\, dx\right| \leq \frac{X^{2j+1}}{j+\frac{1}{2}}, \tag{14}$$

so that $h_n^{(2j)}(0)$ varies within a compact set for each fixed j. Applying Cantor's diagonal argument, we may assume without loss of generality that the sequence $\left\{h_n^{(2j)}(0)\right\}_{n=1}^{\infty}$ converges for every j. Put $c_j = \lim_{n\to\infty} h_n^{(2j)}(0)$ and $h_\infty(t) = \sum_{j=0}^{\infty} \frac{c_j}{(2j)!} t^{2j}$. Then from (14) it follows that h_∞ is an entire function and $h_n(t)$ converges uniformly to $h_\infty(t)$ on compact subsets of $\mathbb{C}$. In particular, $h_\infty(t) \geq 0$ for all $t \in \mathbb{R}$.

Next, for any $g \in \mathcal{C}(X)$ with cosine transform h, we have

$$\begin{aligned}\log(8\pi e^\gamma) - \int_0^\infty \frac{1-g(x)}{2\sinh(x/2)}\, dx + \chi_\Delta(-1)\int_0^\infty \frac{g(x)}{2\cosh(x/2)}\, dx \\ = -\frac{1}{\pi}\int_{\mathbb{R}} \Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+it\right) h(t)\, dt,\end{aligned}$$

where $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$ and $a \in \{0,1\}$ is such that $(-1)^a = \chi_\Delta(-1)$. By Stirling's formula we have $\Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+it\right) = \frac{1}{2}\log(1+|t|) + O(1)$, so that

$$\frac{1}{\pi}\int_{\mathbb{R}} \Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+it\right) h(t)\, dt = \frac{1}{\pi}\int_0^\infty \log(1+t) h(t)\, dt + O(1).$$

Moreover, since $|g(x)| \leq 1$ for all x, we have

$$2\sum_{n=1}^{\infty} \frac{\Lambda(n)\chi_\Delta(n)}{\sqrt{n}} g(\log n) \ll 1,$$

where the implied constant depends only on X.

Returning to our construction, since g_n is a maximizing sequence, we may assume without loss of generality that $l(g_n)$ is bounded below. Together with the above observations, we

thus have that $\frac{1}{\pi}\int_0^\infty \log(1+t)h_n(t)\,dt \le C$ for some constant C. Hence, for any $T>0$, we have

$$\frac{1}{\pi}\int_0^T \log(1+t)h_\infty(t)\,dt = \lim_{n\to\infty}\frac{1}{\pi}\int_0^T \log(1+t)h_n(t)\,dt \le C.$$

Since T is arbitrary and $\log(1+t)h_\infty(t)$ is non-negative, we see that

$$\frac{1}{\pi}\int_0^\infty \log(1+t)h_\infty(t)\,dt \le C. \tag{15}$$

In particular, $h_\infty \in L^1([0,\infty))$, so its cosine transform $g_\infty(x) = \frac{1}{\pi}\int_0^\infty h_\infty(t)\cos(xt)\,dt$ is well-defined and continuous. Moreover, by (15) we have

$$\frac{1}{\pi}\int_T^\infty h_\infty(t)\,dt \le \frac{1}{\pi}\int_0^\infty \frac{\log(1+t)}{\log(1+T)}h_\infty(t)\,dt \le \frac{C}{\log(1+T)},$$

and similarly $\frac{1}{\pi}\int_T^\infty h_n(t)\,dt \le \frac{C}{\log(1+T)}$ for all $T>0$. Therefore,

$$\left|g_n(x) - \frac{1}{\pi}\int_0^T h_n(t)\cos(xt)\,dt\right| \le \frac{C}{\log(1+T)}$$

and

$$\left|g_\infty(x) - \frac{1}{\pi}\int_0^T h_\infty(t)\cos(xt)\,dt\right| \le \frac{C}{\log(1+T)}.$$

Now, let $\varepsilon>0$ be given, and choose $T>0$ large enough that $\frac{C}{\log(1+T)} \le \frac{\varepsilon}{3}$. Further, let $N\in\mathbb{Z}_{\ge 0}$ be such that $n>N$ implies that $|h_n(t)-h_\infty(t)| < \frac{\pi}{3T}\varepsilon$ for all $t\in[0,T]$. Then the above inequalities yield

$$\big|g_n(x)-g_\infty(x)\big| \le \frac{2C}{\log(1+T)} + \left|\frac{1}{\pi}\int_0^T \big(h_n(t)-h_\infty(t)\big)\cos(xt)\,dt\right| < \varepsilon$$

for all $n>N$ and $x\ge 0$. Thus, $g_n(x)$ converges uniformly to $g_\infty(x)$. In particular, g_∞ is supported on $[0,X]$ and satisfies $g_\infty(0)=1$, so it is an element of $\mathcal{C}(X)$.

Finally, let $\delta>0$ be given. By Stirling's formula, there is a number $T_0>0$ such that $\Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\big(\frac{1}{2}+a+it\big) \ge 0$ whenever $|t|\ge T_0$. Moreover, it follows from (15) that the function $\Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\big(\frac{1}{2}+a+it\big)\,h_\infty(t)$ is absolutely integrable, so there exists $T\ge T_0$ such that

$$0 \le \frac{1}{\pi}\int_{\mathbb{R}\setminus[-T,T]} \Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+it\right)h_\infty(t)\,dt < \delta.$$

Therefore,

$$\begin{aligned} l(g_\infty)+\delta &> 2\sum_{n=1}^{\infty}\frac{\Lambda(n)\chi_\Delta(n)}{\sqrt{n}}g_\infty(\log n)-\frac{1}{\pi}\int_{-T}^{T}\Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+it\right)h_\infty(t)\,dt \\ &= \lim_{m\to\infty}\left(2\sum_{n=1}^{\infty}\frac{\Lambda(n)\chi_\Delta(n)}{\sqrt{n}}g_m(\log n)-\frac{1}{\pi}\int_{-T}^{T}\Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+it\right)h_m(t)\,dt\right) \\ &\geq \lim_{m\to\infty}\left(2\sum_{n=1}^{\infty}\frac{\Lambda(n)\chi_\Delta(n)}{\sqrt{n}}g_m(\log n)-\frac{1}{\pi}\int_{\mathbb{R}}\Re\frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+it\right)h_m(t)\,dt\right) \\ &= \sup_{g\in\mathcal{C}(X)} l(g)\geq l(g_\infty). \end{aligned}$$

Since δ is arbitrary, we have $l(g_\infty)=\sup_{g\in\mathcal{C}(X)} l(g)$. □

A.2. **Proof of Prop. 2.2.** We begin with some lemmas.

Lemma A.1. *For $\nu>0$, define*

$$f_\nu(x)=\begin{cases}\left(1-x^2\right)^\nu & \text{if } |x|<1,\\ 0 & \text{otherwise,}\end{cases}$$

$$g_{\nu,X}(x)=\frac{\Gamma(\frac{3}{2}+2\nu)}{\sqrt{\pi}\Gamma(1+2\nu)}\int_{\mathbb{R}}f_\nu(y)f_\nu\left(\frac{2x}{X}-y\right)dy \quad \text{for } x\geq 0,$$

and $h_{\nu,X}(t)=2\int_0^\infty g_{\nu,X}(x)\cos(tx)\,dx$. Then $g_{\nu,X}(0)=1$, $g_{\nu,X}$ is supported on $[0,X]$, $h_{\nu,X}$ is non-negative and satisfies

$$h_{\nu,X}(t)\ll_\varepsilon \nu^{-1/2}e^{-2\nu}X\left|\frac{4\nu}{Xt}\right|^{2\nu+2} \qquad \text{uniformly for } \left|\frac{Xt}{4}\right|\geq\nu\geq\varepsilon,$$

for any fixed $\varepsilon>0$.

Proof. Using the Poisson representation for the J-Bessel function, we derive

$$\int_{\mathbb{R}}f_\nu(x)e^{itx}\,dx=2\Gamma(1+\nu)j_\nu(|t|)\left|\frac{2}{t}\right|^\nu,$$

where $j_\nu(u)=\sqrt{\frac{\pi}{2u}}J_{\nu+1/2}(u)$ is the spherical Bessel function with parameter ν. From this and the limit $j_\nu(u)(\frac{2}{u})^\nu\to\frac{\sqrt{\pi}}{2\Gamma(\frac{3}{2}+\nu)}$ as $u\to 0^+$, we derive

$$\int_{\mathbb{R}}f_\nu(x)^2\,dx=\int_{\mathbb{R}}f_{2\nu}(x)\,dx=\frac{\sqrt{\pi}\Gamma(1+2\nu)}{\Gamma(\frac{3}{2}+2\nu)},$$

so that $g_{\nu,X}(0)=1$. Therefore, we have

$$h_{\nu,X}(t)=\frac{2X}{\sqrt{\pi}}\frac{\Gamma(\frac{3}{2}+2\nu)\Gamma(1+\nu)^2}{\Gamma(1+2\nu)}j_\nu\left(\frac{X|t|}{2}\right)^2\left|\frac{4}{Xt}\right|^{2\nu}.$$

It follows from [15, Thm. 2] that the function $uj_\nu(u)$ is bounded in the region $\{(\nu,u):u\geq 2\nu\geq 0\}$. This combined with Stirling's formula gives the estimate. □

Lemma A.2. *For $\alpha > 0$, set*

$$H_\alpha(t) = \frac{\alpha}{2}\operatorname{sinc}^2\left(\frac{\alpha t}{2}\right) + \frac{\alpha}{4}\left[\operatorname{sinc}^2\left(\frac{\alpha t - \pi}{2}\right) + \operatorname{sinc}^2\left(\frac{\alpha t + \pi}{2}\right)\right]$$

and $G_\alpha(x) = \frac{1}{\pi}\int_0^\infty H_\alpha(t)\cos(tx)\,dt$ for $x \geq 0$. Then G_α is supported on $[0, \alpha]$, $G_\alpha(0) = 1$, and $H_\alpha(t) \geq \frac{2\alpha}{\max((\alpha t)^2, \pi^2/3)}$ for $t \in \mathbb{R}$.

Proof. A straightforward calculation gives

$$G_\alpha(x) = \max\Big(0, 1 - \frac{x}{\alpha}\Big)\cos^2\Big(\frac{\pi x}{2\alpha}\Big), \tag{16}$$

which yields the stated properties of G_α. As for H_α, by rescaling, it suffices to prove the bound for $\alpha = 1$. A calculation shows that

$$H_1(t) = \frac{2}{t^2} + 2\left(\frac{\pi\cos(t/2)}{t(t^2 - \pi^2)}\right)^2 (3t^2 - \pi^2),$$

so that $H_1(t) \geq 2t^{-2}$ for $|t| \geq \pi/\sqrt{3}$. On the other hand, graphing the function verifies that $H_1(t) \geq 6/\pi^2$ for $|t| \leq \pi/\sqrt{3}$. □

Now, turning to Prop. 2.2, first note that $|q\Delta| \geq 3$. We take $h(t) = h_{\nu,X}(t)$, where $\nu = \frac{\delta X}{4} \geq \frac{1}{2}\log\log 3 > 0$. Applying Lemma A.1 with $\varepsilon = \frac{1}{2}\log\log 3$ and Lemma A.2 with $\alpha = 2\log\log|q\Delta|$, for $t \geq \delta$ we have

$$\begin{aligned} h(t) &\ll \nu^{-1/2}e^{-2\nu}X\left(\frac{4\nu}{Xt}\right)^{2\nu+2} \leq X(\delta X)^{-1/2}e^{-\delta X/2}\alpha^{-1}\max\left((\alpha\delta)^2, \frac{\pi^2}{3}\right)H_\alpha(t) \\ &\leq \frac{e^{-A}X}{(2\log\log|q\Delta|)^{3/2}}\frac{\max((2\delta\log\log|q\Delta|)^2, \pi^2/3)}{\log|q\Delta|}H_\alpha(t). \end{aligned}$$

Since $\gamma_j(q\Delta) \geq \delta$ for every j, this yields

$$\sum_{j=1}^\infty h(\gamma_j(q\Delta)) \ll \frac{e^{-A}X}{(\log\log|q\Delta|)^{3/2}}\frac{\max((2\delta\log\log|q\Delta|)^2, \pi^2/3)}{\log|q\Delta|}\sum_{j=1}^\infty H_\alpha(\gamma_j(q\Delta)).$$

We estimate the latter sum by plugging back into the explicit formula (3), with Δ replaced by $q\Delta$. A calculation with the prime number theorem using (16) shows that

$$\sum_{n=1}^\infty \frac{\Lambda(n)}{\sqrt{n}}G_\alpha(\log n) \ll \frac{e^{\alpha/2}}{\alpha^3},$$

and it is not hard to see that

$$\log(8\pi e^\gamma) - \int_0^\infty \frac{1 - G_\alpha(x)}{2\sinh(x/2)}\,dx + \chi_{q\Delta}(-1)\int_0^\infty \frac{G_\alpha(x)}{2\cosh(x/2)}\,dx = O(1)$$

uniformly for $\alpha \geq 2\log\log 3$. Thus, $\sum_{j=1}^\infty H_\alpha(\gamma_j(q\Delta)) \ll \log|q\Delta|$.

Finally, by [24, Thm. 11], under GRH we have $\delta \ll 1/\log\log|q\Delta|$. This yields (7). □

A.3. **Proof of Prop. 2.3.** We may assume without loss of generality that N is odd, not a perfect square, and satisfies $N \geq \exp(\exp(C^{2/3}))$, where $C > 0$ is the implied constant in (7).[18] Thus, if we set $d = (-1)^{\frac{N-1}{2}} N$ then $d = \Delta \ell^2$ for some $1 \neq \Delta \in \mathcal{F}$ and $\ell \in \mathbb{Z}_{>0}$.

Let $Q \geq 3$ be an integer parameter to be specified later. Set $\nu = \frac{1}{2} \log\log N + \frac{1}{12 \log\log N}$, $X = 4\nu \log Q$, and let T be an integer in the interval $[e^X - 1, e^X + 1)$. (Note that to find such a T, it suffices to compute e^X to within $\pm\frac{1}{2}$.) We let q run through all elements of $\mathcal{F}$ with $|q| \leq Q$ and evaluate the lower bound (6) using the test function $g = g_{\nu,X}$, in the notation of Lemma A.1.

Since $g_{\nu,X}(\log n) = 0$ for $n > T$, it is enough to consider the terms of the sum for $n \leq T$. As described in §1.2, since Δ is unknown to us, we compute $\chi_d(n)$ in place of $\chi_\Delta(n)$. If for any prime value of n we find a zero value of $\chi_d(n)$, we check to see if $n^2 | N$ and exit with this square factor if so; otherwise $\chi_\Delta(n) = \chi_d(n)$. In particular, while computing (6) for $q = 1$, we evaluate $\chi_\Delta(n)$ for all primes $n \leq T$. If $(T+1)^3 > N$ then this alone yields enough information to determine whether N is squarefree. Hence, we may assume without loss of generality that $X \leq \frac{1}{3} \log N$, so that $\log Q \leq \frac{\log N}{12\nu}$.

Note that if we set $A = 2\nu - \log\log|q\Delta|$ and $\delta = \frac{1}{\log Q}$ then $g_{\nu,X}$ is precisely the test function exhibited in the proof of Prop. 2.2. Using the bound $|q\Delta| \leq QN \leq N^{1+\frac{1}{12\nu}}$, we derive the inequality $1 - e^{-A} > \frac{1}{72}(\log\log N)^{-2} > 0$. Thus, Prop. 2.2 shows that if $|\Delta| = N$ (which holds when N is squarefree) and $\gamma_1(q\Delta) \geq \frac{1}{\log Q}$ then

$$\sum_{j=1}^{\infty} h_{\nu,X}(\gamma_j(q\Delta)) < \left(1 - \frac{1}{72(\log\log N)^2}\right) \frac{CX}{(\log\log(qN))^{3/2}} \leq \left(1 - \frac{1}{72(\log\log N)^2}\right) X.$$

Therefore, if we evaluate (6) to within $\pm\frac{X}{72(\log\log N)^2}$, we will have proven that $|\Delta| > Ne^{-2X}$, so that $\ell \leq T$. Having already determined all prime factors of N up to T, we will thus have found a proof that N is squarefree.

Now, since the value of θ^* is unknown to us, we cannot say in advance what value of Q will suffice. In our algorithm, we therefore apply the above procedure iteratively with $Q = 2^k$ for $k = 2, 3, 4, \ldots$ until we find either a square factor or a proof that N is squarefree. As noted above, the algorithm must eventually terminate. If it turns out that $\theta^* = 1$ or if N has a square factor then the algorithm becomes a rather inefficient version of trial division, which nevertheless runs in polynomial time in N; in particular, the $O(\exp[(\log N)^{1+o(1)}])$ running time estimate holds. Henceforth we will assume that $\theta^* < 1$ and that the input N is squarefree.

Fix $\varepsilon \in (0, 1 - \theta^*)$. From the definition of θ^* it follows that $\eta_\infty(\theta^* + \varepsilon) \leq \theta^*$. Thus, there exists $N_0(\varepsilon) \in \mathbb{Z}_{>0}$ such that $\eta_\Delta(\theta^* + \varepsilon) < \theta^* + \varepsilon$ whenever $|\Delta| = N \geq N_0(\varepsilon)$. Let us assume that $N \geq N_0(\varepsilon)$. Then once $\frac{\log\log Q}{\log\log N} \geq \theta^* + \varepsilon$, there must be a q with $(q, \Delta) = 1$ and $|q| \leq Q$ such that $\gamma_1(q\Delta) > \frac{1}{\log Q}$.

It is straightforward to see that all of the floating point operations required to compute (6) for every $|q| \leq Q$ to the precision described above may be carried out in time $O(Q^{1+4\nu} \log^c N)$ for some $c > 0$. Since we choose values of Q from a geometric progression, the total running

[18] If C is at all large then this rules out every N of practical size; we could deal with this instead by increasing A in Prop. 2.2 by a constant, but as we are only interested in the theoretical result, we make this assumption for convenience.

time is dominated by that of the final iteration. In the worst case, it might be that the smallest Q for which $\frac{\log\log Q}{\log\log N} \geq \theta^* + \varepsilon$ is $2^k + 1$ for some k, and thus our final choice of $Q = 2^{k+1}$ would be too large by roughly a factor of 2. Thus, $\log Q \leq (\log N)^{\theta^*+\varepsilon} + \log 2$, so that $Q^{1+4\nu}\log^c N \ll \exp\big[(\log N)^{\theta^*+\varepsilon}(1+4\nu)\big](\log N)^{c+\log 4}$. Since $1 + 4\nu \ll \log\log N$ and ε may be chosen arbitrarily small (assuming only that $N \geq N_0(\varepsilon)$), the running time is thus $O\big(\exp\big[(\log N)^{\theta^*+o(1)}\big]\big)$, as required. □

A.4. **Proof of Prop. 3.3.**

Lemma A.3. *For each $N \geq 1$, and each $s \in (0,\pi)$, we have*

(17)
$$1 \leq \frac{\mathbb{P}_{USp(2N)}\big(\theta_1 > s\big)}{\cos(s/2)^{N(2N+1)}} \leq \frac{1}{2}\left(1 + \frac{\sin(s/2)}{\sqrt{2}}\right)^{2N+1} + \frac{1}{2}\left(1 - \frac{\sin(s/2)}{\sqrt{2}}\right)^{2N+1} \leq \exp\left(\frac{Ns}{\sqrt{2}}\right).$$

In particular, we have

$$\log \mathbb{P}_{USp(2N)}\big(\theta_1 > s\big) = \big[2 + O\big((Ns)^{-1}\big)\big]N^2 \log\cos(s/2), \tag{18}$$

uniformly for $s \in (0,\pi)$.

Proof. The Weyl integration formula on $USp(2N)$ gives

$$\begin{aligned}\mathbb{P}_{USp(2N)}\big(\theta_1 > s\big) &= \int_{USp(2N)} 1_{\theta_1 > s}\, d\mathbb{P}_{USp(2N)} \\ &= \frac{2^{N^2}}{\pi^N\, N!}\int_{(s,\pi]^N} \prod_{1\leq j<k\leq N} (\cos\phi_k - \cos\phi_j)^2 \prod_{1\leq j\leq N} (\sin^2\phi_j)\, d\phi_1\cdots d\phi_N.\end{aligned}$$

We proceed along the lines of the proof of [13, proposition 6.10.1].[19] Applying the change of variable $\phi_j = 2\tau_j$, the trig identities $\sin(2\tau_j) = 2\sin\tau_j\cos\tau_j$ and $\cos(2\tau_j) = 2\cos^2\tau_j - 1$, and, last, the substitution $w_j = \cos^2\tau_j$, we obtain $\mathbb{P}_{USp(2N)}\big(\theta_1 > s\big) = I_N(\cos^2(s/2))$, where

$$I_N(\lambda) := \frac{C(N)}{N!}\int_{[0,\lambda)^N} \prod_{1\leq j<k\leq N} (w_k - w_j)^2 \prod_{1\leq j\leq N} \sqrt{w_j(1-w_j)}\, dw_1\cdots dw_N,$$

and $C(N) := 2^{2N^2+N}/\pi^N$. The change of variable $w_j = \lambda x_j$ thus yields

$$I_N(\lambda) = \lambda^{N^2+N/2}\frac{C(N)}{N!}\int_{[0,1)^N} \prod_{1\leq j<k\leq N} (x_k - x_j)^2 \prod_{1\leq j\leq N} \sqrt{x_j(1-\lambda x_j)}\, dx_1\cdots dx_N. \tag{19}$$

Since $\sqrt{1-\lambda x_j} \geq \sqrt{1-x_j}$ for $0 \leq \lambda \leq 1$, we have $I_N(\lambda) \geq \lambda^{N^2+N/2} I_N(1)$. The lower bound follows on observing that $I_N(1) = \mathbb{P}_{USp(2N)}\big(\theta_1 > 0\big) = 1$.

By comparing the joint probability density functions of eigenphases in $USp(2N)$ and $SO(2N)$, one easily obtains the rough upper bound $\mathbb{P}_{USp(2N)}\big(\theta_1 > s\big) \leq 4^N \mathbb{P}_{SO(2N)}\big(\theta_1 > s\big)$. Combined with the estimate $\mathbb{P}_{SO(2N)}\big(\theta_1 > s\big) \leq \cos(s/2)^{2N^2-N}$ from [13, proposition 6.10.1] and the lower bound $I_N(\lambda) \geq \lambda^{N^2+N/2}$, this yields the asymptotic (18) in the range $\beta \in (1,2)$.

[19] $\mathbb{P}_{G(N)}\big(\theta_1 > s\big)$ is eigen$(0, s, G(N))$ in the notation of [13].

To derive the upper bound, and consequently the asymptotic, in the full range, we apply the first-order inequality

$$\begin{aligned}\sqrt{1-\lambda x_j} &= \sqrt{1-x_j}\sqrt{1+\frac{(1-\lambda)x_j}{1-x_j}} \le \sqrt{1-x_j}\left(1+\frac{(1-\lambda)x_j}{2(1-x_j)}\right)\\ &= \frac{1+\lambda}{2}\sqrt{1-x_j}\left(1+\frac{1-\lambda}{1+\lambda}\frac{1}{1-x_j}\right),\end{aligned}$$

to get

$$\begin{aligned}\prod_{j=1}^{N}\sqrt{1-\lambda x_j} &\le \left(\frac{1+\lambda}{2}\right)^N \prod_{j=1}^{N}\sqrt{1-x_j}\cdot \sum_{J\subset\{1,\dots,N\}}\left(\frac{1-\lambda}{1+\lambda}\right)^{\#J}\prod_{j\in J}(1-x_j)^{-1}\\ &= \left(\frac{1+\lambda}{2}\right)^N \prod_{j=1}^{N}\frac{1}{\sqrt{1-x_j}}\cdot \sum_{J\subset\{1,\dots,N\}}\left(\frac{1-\lambda}{1+\lambda}\right)^{\#J}\prod_{j\in\{1,\dots,N\}\setminus J}(1-x_j),\end{aligned}$$

where J runs through all subsets of $\{1,\dots,N\}$. We insert this into (19) and permute the variables so that $\{1,\dots,N\}\setminus J$ is mapped to $\{1,\dots,N-\#J\}$. Collecting the terms with a common value of $\#J$, we have

$$\begin{aligned}\frac{I_N(\lambda)}{\lambda^{N^2+N/2}} \le \frac{C(N)}{N!}\left(\frac{1+\lambda}{2}\right)^N \sum_{m=0}^{N}\binom{N}{m}\left(\frac{1-\lambda}{1+\lambda}\right)^m\\ \cdot\int_{[0,1)^N}\prod_{1\le j<k\le N}(x_k-x_j)^2\prod_{j=1}^{N}\sqrt{\frac{x_j}{1-x_j}}\prod_{j=1}^{N-m}(1-x_j)\,dx_1\cdots dx_N.\end{aligned}$$

By Aomoto's formula [20, (17.1.6)], the integral may be written in the form

$$K_N\prod_{j=1}^{N-m}\frac{\frac12+N-j}{1+2N-j} = K_N\frac{\Gamma(N+\frac12)}{\Gamma(m+\frac12)}\frac{\Gamma(1+N+m)}{\Gamma(1+2N)},$$

where K_N (a Selberg integral, see [20, (17.1.3)]) is independent of m. When $m=0$, the integral is easily recognized as the one occurring in (19) with $\lambda=1$, so that

$$1 = I_N(1) = \frac{C(N)K_N}{N!}\frac{\Gamma(N+\frac12)}{\Gamma(\frac12)}\frac{\Gamma(1+N)}{\Gamma(1+2N)}.$$

Solving for $C(N)K_N/N!$ and substituting back into the above, we have

$$\begin{aligned}\frac{I_N(\lambda)}{\lambda^{N^2+N/2}} &\le \left(\frac{1+\lambda}{2}\right)^N\sum_{m=0}^{N}\binom{N}{m}\left(\frac{1-\lambda}{1+\lambda}\right)^m\frac{\Gamma(\frac12)}{\Gamma(m+\frac12)}\frac{\Gamma(1+N+m)}{\Gamma(1+N)}\\ &= \left(\frac{1+\lambda}{2}\right)^N\sum_{m=0}^{N}\binom{N+m}{2m}\left(4\frac{1-\lambda}{1+\lambda}\right)^m.\end{aligned}$$

The last sum is known as a Morgan-Voyce polynomial; it is closely related to the Chebyshev polynomials, and may be evaluated in closed form. Precisely, if t is such that $\cosh t = \sqrt{\frac{2}{1+\lambda}}$,

then it follows from [31, (11b)] that the last line is

$$\frac{\cosh((2N+1)t)}{\cosh^{2N+1}t} = \frac{1}{2}\left(1+\sqrt{\frac{1-\lambda}{2}}\right)^{2N+1} + \frac{1}{2}\left(1-\sqrt{\frac{1-\lambda}{2}}\right)^{2N+1}.$$

The upper bound follows on putting $\lambda = \cos^2(s/2)$ and noting that

$$\frac{1}{2}\left(1+\frac{\sin(s/2)}{\sqrt{2}}\right)^{2N+1} + \frac{1}{2}\left(1-\frac{\sin(s/2)}{\sqrt{2}}\right)^{2N+1} \le \exp\left(\frac{Ns}{\sqrt{2}}\right).$$

Combining this with the lower bound and the inequality $|\log\cos(s/2)| \ge s^2/8$, we get the estimate

$$\log \mathbb{P}_{USp(2N)}(\theta_1 > s) = N(2N+1)\log\cos(s/2) + O(Ns) = \big[2 + O\big((Ns)^{-1}\big)\big]N^2\log\cos(s/2).$$

□

Turning to the proof of Prop. 3.3, by definition of $\mathbb{P}_{USp(2N)}(\max_{1\le m\le M}\theta_1(m) \le s)$, we have, for each $s \in [0,\pi]$,

$$\mathbb{P}_{USp(2N)}\big(\max_{1\le m\le M}\theta_1(m) \le s\big) = \mathbb{P}_{USp(2N)}\big(\theta_1 \le s\big)^M. \tag{20}$$

Suppose $\varepsilon < 4$ and $s \in [s^-_{\varepsilon,\beta}(N), s^+_{\varepsilon,\beta}(N)]$. Then $s \to 0$ as $N \to \infty$ (since $\beta < 2$ by assumption). Further, by Lemma A.3 we have $\mathbb{P}_{USp(2N)}(\theta_1 > s) \to 0$ as $N \to \infty$ (since $\beta > 0$ by assumption). Using (20), and Lemma A.3 again, yields

(21)

$$\begin{aligned}
\log \mathbb{P}_{USp(2N)}\big(\max_{1\le m\le M}\theta_1(m) \le s\big) &= M\log\big(1 - \mathbb{P}_{USp(2N)}(\theta_1 > s)\big)\\
&= -M\,\mathbb{P}_{USp(2N)}\big(\theta_1 > s\big)(1+o(1))\\
&= -\exp\big((2N)^\beta + [2+O((Ns)^{-1})]N^2\log\cos(s/2) + o(1)\big)\\
&= -\exp\big((2N)^\beta - (1+o(1))(Ns/2)^2 + o(1)\big),
\end{aligned}$$

where we used that $\log\cos(s/2) \sim -s^2/8$ in the last line. So if $s = s^-_{\varepsilon,\beta}(N) = (4-\varepsilon)(2N)^{\beta/2-1}$, then

$$\begin{aligned}
\log \mathbb{P}_{USp(2N)}\big(\max_{1\le m\le M}\theta_1(m) \le s\big) &= -\exp\big((2N)^\beta - (1+o(1))(1-\varepsilon/4)^2(2N)^\beta + o(1)\big)\\
&= -\exp\big((1+o(1))(2N)^\beta(8\varepsilon-\varepsilon^2)/16 + o(1)\big) \to -\infty,
\end{aligned} \tag{22}$$

as $N \to \infty$, provided that $0 < \varepsilon < 4$, which ensures that $(2N)^\beta(8\varepsilon-\varepsilon^2)/16 \to \infty$. If $\varepsilon \ge 4$ then clearly the result still holds. Therefore, for each $\varepsilon > 0$, we have $\mathbb{P}_{USp(2N)}\big(s^-_{\varepsilon,\beta}(N) < \max_{1\le m\le M}\theta_1(m)\big) \to 1$, as claimed.

Similarly, if $s = s^+_{\varepsilon,\beta}(N) = (4+\varepsilon)(2N)^{\beta/2-1}$, then

$$\begin{aligned}
\log \mathbb{P}_{USp(2N)}\big(\max_{1\le m\le M}\theta_1(m) \le s\big) &= -\exp\big((2N)^\beta - (1+o(1))(1+\varepsilon/4)^2(2N)^\beta + o(1)\big)\\
&= -\exp\big(-(1+o(1))(2N)^\beta(8\varepsilon+\varepsilon^2)/16 + o(1)\big),
\end{aligned}$$

which tends to 0 with N. Therefore, $\mathbb{P}_{USp(2N)}\big(\max_{1\le m\le M}\theta_1(m) \le s^+_{\varepsilon,\beta}(N)\big) \to 1$. □

A.5. **Proof of Prop. 3.4.** We make use of the main results in [7] and [16], formulated in Lemma A.4 here.

Lemma A.4. *For $\delta > 0$ fixed, there exists a (large) positive constant s_0 such that*

$$\log \mathbb{P}_{U(N)}\big(\theta_1 > 2s\big) = N^2 \log\cos(s/2) - \frac{1}{4}\log\left(N\sin(s/2)\right) + c_0 + O\left(1/(N\sin(s/2))\right)$$

$$\frac{d}{ds}\log \mathbb{P}_{U(N)}\big(\theta_1 > 2s\big) = -\frac{N^2}{2}\tan(s/2) - \frac{1}{8}\cot(s/2) + O(1/(N\sin^2(s/2))),$$

for all $n > s_0$ and $2s_0/N \le s \le \pi - \delta$, where c_0 is an explicit constant.

Proof. Clearly, $\mathbb{P}_{U(N)}\big(\theta_1 > 2s\big) = \int_{U(N)} 1_{\theta_1 > 2s}\, d\mathbb{P}_{U(N)}$. By the rotational invariance of $\mathbb{P}_{U(N)}$, this is the same as $\int_{U(N)} 1_{\theta_1,\dots,\theta_N \notin [0,s]\cup[2\pi-s,2\pi]}\, d\mathbb{P}_{U(N)}$. Further, by [4, Lemma 2] and the Weyl integration formula on $U(N)$, this is the Toeplitz determinant $\det_{N\times N}(\int_s^{2\pi-s} e^{i(j-k)\theta}\, d\theta/(2\pi))$, for which the relevant asymptotics are supplied by formulas (8) and (12) in [7]. □

Remark. Lemma 6.8.3 in [13] furnishes the following interesting factorization of the gap probabilities: $\mathbb{P}_{U(2N+1)}\big(\theta_1 > 2s\big) = \mathbb{P}_{SO(2N+2)}\big(\theta_1 > s\big)\, \mathbb{P}_{USp(2N)}\big(\theta_1 > s\big)$. So, as a direct consequence of Lemmas A.3 and A.4, we obtain $\log \mathbb{P}_{SO(2N)}\big(\theta_1 > s\big) = (2+o(1))N^2 \log\cos(s/2)$, provided that $Ns \to \infty$ as $N \to \infty$. Note that the machinery of orthogonal polynomials supplies general, but involved, methods to derive precise asymptotics for determinant expressions of gap probabilities like those in (23); e.g. see [3] and [8]. In the case of $USp(2N)$, for example, the gap probability can be expressed as Toeplitz+Hankel determinant, or by appealing to (19), as a Hankel determinant.

To prove the proposition, first note[20]

$$\frac{N}{2\pi}\mathbb{P}_{U(N)}(\theta_2 - \theta_1 > u) = -\frac{1}{2}\frac{d}{ds}\mathbb{P}_{U(N)}\big(\theta_1 > 2s\big)\Big|_{s=u/2}. \tag{23}$$

Therefore,

$$\mathbb{P}_{U(N)}(\theta_2 - \theta_1 > u) = -\frac{\pi}{N}\left(\frac{d}{ds}\log \mathbb{P}_{U(N)}\big(\theta_1 > 2s\big)\right)\Big|_{s=u/2} \mathbb{P}_{U(N)}\big(\theta_1 > u\big). \tag{24}$$

It follows from Lemma A.4 that as $N \to \infty$, and uniformly for $N^{\nu_1 - 1} \le u \le 2\pi - \delta$ (for any small constant $\nu_1 > 0$ we wish), that the term controlling the behavior in (24) is $\mathbb{P}_{U(N)}\big(\theta_1 > u\big)$. Explicitly,

$$\mathbb{P}_{U(N)}(\theta_2 - \theta_1 > u) = \exp((1+o(1))N^2 \log\cos(u/4))\,.$$

If we further require $u \le N^{-\nu_2}$ (for any small constant $\nu_2 > 0$ we wish), then $u \to 0$ as $N \to \infty$ and $\log\cos(u/4) \sim -u^2/32$. Therefore,

$$\mathbb{P}_{U(N)}(\tfrac{1}{2}[\theta_2 - \theta_1] > u) = \exp(-(1+o(1))N^2u^2/8) \tag{25}$$

uniformly for $N^{\nu_1 - 1} \le 2u \le N^{-\nu_2}$. Thus, by a similar calculation to (21), we obtain the result for $\max_{1\le j\le M_\beta(N)} \frac{1}{2}[\theta_2(m) - \theta_1(m)]$. If, in addition, one maximizes over all the spacings

[20] See Lemma 3.1 in [2], but note it is missing a factor of 2π.

in each matrix, thus replacing $\frac{1}{2}[\theta_2(m)-\theta_1(m)]$ by $\max_{1\le j\le N}\frac{1}{2}[\theta_{j+1}(m)-\theta_j(m)]$, then the same result holds. For clearly

$$\mathbb{P}_{U(N)}\big(u^-_{\varepsilon,\beta}(N) < \max_{1\le m\le M_\beta(N)}\tfrac{1}{2}[\theta_2-\theta_1]\big) \le \mathbb{P}_{U(N)}\big(u^-_{\varepsilon,\beta}(N) < \max_{1\le m\le M_\beta(N)}\max_{1\le j\le N}\tfrac{1}{2}[\theta_{j+1}(m)-\theta_j(m)]\big).$$

Hence, if the left-hand tends to 1 then the right-hand side does as well. In the opposite direction, we have

$$N\,\mathbb{P}_{U(N)}\big(\tfrac{1}{2}[\theta_2-\theta_1] > u^+_{\varepsilon,\beta}(N)\big) = o(1),$$

as $N\to\infty$, by virtue of (25). Thus,

$$\begin{aligned}\log\mathbb{P}_{U(N)}\big(\max_{1\le m\le M_\beta(N)}\max_{1\le j\le N}\tfrac{1}{2}[\theta_{j+1}(m)-\theta_j(m)] \le u^+_{\varepsilon,\beta}(N)\big)&\\ = M\log\mathbb{P}_{U(N)}\big(\max_{1\le j\le N}\tfrac{1}{2}[\theta_{j+1}(m)-\theta_j(m)] \le u^+_{\varepsilon,\beta}(N)\big)&\\ \ge M\log\big(1-N\mathbb{P}_{U(N)}\big(\tfrac{1}{2}[\theta_2-\theta_1] > u^+_{\varepsilon,\beta}(N)\big)\big)&\\ = -MN\,\mathbb{P}_{U(N)}\big(\tfrac{1}{2}[\theta_2-\theta_1] > u^+_{\varepsilon,\beta}(N)\big)(1+o(1)),&\end{aligned}$$

where we used the rotational invariance of $\mathbb{P}_{U(N)}$ in the third line. Hence maximizing over $1\le j\le N$ does not change our previous calculation more than does increasing the number of samples by a factor of N, which affects lower order terms only. Thus, as before, $-MN\,\mathbb{P}_{U(N)}\big(\frac{1}{2}[\theta_2-\theta_1] > u^+_{\varepsilon,\beta}(N)\big)\to 0$, and so $\mathbb{P}_{U(N)}\big(\max_{1\le m\le M_\beta(N)}\max_{1\le j\le N}\frac{1}{2}[\theta_{j+1}(m)-\theta_j(m)] \le u^+_{\varepsilon,\beta}(N)\big)\to 1$. □

A.6. **Proof of Prop. 4.2.** By the growth estimate for the divisor m, there is a Hadamard product $F(z)$ which is entire of finite order, even, satisfies $\operatorname{ord}_{z=\gamma}F(z)=m(\gamma)$ for all $\gamma\in S$ and does not vanish outside of S. Moreover, $\frac{F'}{F}(z)$ has at most polynomial growth in horizontal strips outside of S, so that $\frac{F'}{F}(z)h(z)$ decays rapidly in such strips for any h as in the statement of the proposition. By the argument principle, for any $c>1/2$ we have

$$\begin{aligned}\sum_{\gamma\in S}m(\gamma)h(\gamma) &= \frac{1}{2\pi i}\int_{\Im(z)=-c}\frac{F'}{F}(z)h(z)\,dz - \frac{1}{2\pi i}\int_{\Im(z)=c}\frac{F'}{F}(z)h(z)\,dz\\ &= \frac{1}{\pi i}\int_{\Im(z)=-c}\frac{F'}{F}(z)h(z)\,dz.\end{aligned}$$

Next, let $a\in\{0,1\}$ be such that $(-1)^a=\operatorname{sgn} d$, and define

$$\Lambda(s)=|d|^{s/2}\Gamma_{\mathbb{R}}(s+a)\exp\left(\sum_{n=2}^{\infty}\frac{c_n}{\log n}n^{\frac{1}{2}-s}\right)\quad\text{and}\quad\Phi(z)=\Lambda(\frac{1}{2}+iz).$$

By the estimate for c_n, Φ is analytic for $\Im(z)<-1$, where it satisfies

$$-i\frac{\Phi'}{\Phi}(z)=\frac{1}{2}\log|d|+\frac{\Gamma'_{\mathbb{R}}}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+iz\right)-\sum_{n=2}^{\infty}c_n n^{-iz}.$$

Thus, for any $c > 1$ we have

$$\begin{aligned}
&\frac{1}{\pi i}\int_{\Im(z)=-c} \frac{\Phi'}{\Phi}(z)h(z)\,dz \\
&\quad = g(0)\log|d| + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\Gamma_{\mathbb{R}}'}{\Gamma_{\mathbb{R}}}\left(\frac{1}{2}+a+it\right)h(t)\,dt - 2\sum_{n=2}^{\infty} c_n g(\log n) \\
&\quad = \sum_{\gamma\in S} m(\gamma)h(\gamma) = \frac{1}{\pi i}\int_{\Im(z)=-c} \frac{F'}{F}(z)h(z)\,dz.
\end{aligned}$$

Let us now set $f(z) = \frac{F'}{F}(z) - \frac{\Phi'}{\Phi}(z)$ for $\Im(z) < -1$. By the above, we see that $\frac{1}{\pi}\int_{\Im(z)=-c} f(z)h(z)\,dz = 0$ for every $c > 1$ and every suitable choice of test function h. Fix one choice of h and consider the Fourier transform

$$u(x) = \frac{1}{2\pi}\int_{\Im(z)=-c} f(z)h(z)e^{-ixz}\,dz.$$

Note that since $f(z)h(z)$ is holomorphic for $\Im(z) < -1$ and of rapid decay in horizontal strips, $u(x)$ does not depend on c. Further, for any fixed $x \in \mathbb{R}$, $h(z)\cos(xz)$ is also a suitable test function, so we have

$$u(x) + u(-x) = \frac{1}{\pi}\int_{\Im(z)=-c} f(z)h(z)\cos(xz)\,dz = 0,$$

i.e. u is an odd function of x. Combining this with the trivial estimate $u(x) \ll_{c,h} e^{-cx}$, we get $u(x) \ll_{c,h} e^{-c|x|}$.

Using this estimate for some $c > 1$ together with the Fourier inversion formula

$$f(z)h(z) = \int_{-\infty}^{\infty} u(x)e^{ixz}\,dx,$$

we see that $f(z)h(z)$ continues to an entire function and is odd. Since h is arbitrary, it follows from a suitable approximation argument that f continues to an odd entire function with at most polynomial growth in horizontal strips. Recalling the definition of f and integrating, we see that $\Phi(z)$ continues to an entire function of finite order satisfying $\Phi(z) = \epsilon\Phi(-z)$ for some $\epsilon \in \{\pm 1\}$, $\operatorname{ord}_{z=\gamma}\Phi(z) = m(\gamma)$ for every $\gamma \in S$, and $\Phi(z) \neq 0$ for $z \notin S$.

The remaining statements now essentially follow from the converse theorem for degree 1 elements of the Selberg class [12], except that the proof given there assumes that the Dirichlet series $L(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ defined by $L(s) = \exp\Big(\sum_{n=2}^{\infty} \frac{c_n}{\log n} n^{\frac{1}{2}-s}\Big)$ converges absolutely for $\Re(s) > 1$, which we only know to be true for $\Re(s) > \frac{3}{2}$. That assumption is not necessary, however, and for the sake of completeness we sketch a simplified proof following the method of [29].

First note that the symmetry of Φ is equivalent to the functional equation $\Lambda(s) = \epsilon\Lambda(1-s)$. Next, for any $\alpha, y > 0$ we have

$$\begin{aligned}
2\sum_{n=1}^{\infty} a_n e(n\alpha)e^{-2\pi n y} &= \frac{1}{2\pi i}\int_{\Re(s)=2} L(s)\Gamma_{\mathbb{C}}(s)(y-i\alpha)^{-s}\,ds \\
&= \frac{1}{2\pi i}\int_{\Re(s)=2} \Lambda(s)\Gamma_{\mathbb{R}}(s+1-a)\big[\sqrt{|d|}(y-i\alpha)\big]^{-s}\,ds,
\end{aligned}$$

where for any z with positive real part we define $z^{-s} = \exp(-s \log z)$ using the principal branch of the logarithm.

By the Phragmén–Lindelöf theorem, the integrand decays rapidly in vertical strips, so we may shift the contour to $\Re(s) = -3/4$ and apply the functional equation to obtain

$$\begin{aligned} 2\sum_{n=1}^{\infty} a_n e(n\alpha) e^{-2\pi n y} - (1-(-1)^a)\Lambda(0) \\ &= \frac{1}{2\pi i}\int_{\Re(s)=-3/4} \Lambda(s)\Gamma_{\mathbb{R}}(s+1-a)\big[\sqrt{|d|}(y-i\alpha)\big]^{-s}\,ds \\ &= \frac{\epsilon}{2\pi i}\int_{\Re(s)=7/4} \Lambda(s)\Gamma_{\mathbb{R}}(2-a-s)\big[\sqrt{|d|}(y-i\alpha)\big]^{s-1}\,ds. \end{aligned}$$

Expanding $\Lambda(s)$ as $|d|^{s/2}\Gamma_{\mathbb{R}}(s+a)\sum_{n=1}^{\infty} a_n n^{-s}$ and using the identity $\Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(2-s) = \csc(\pi s/2)$, we get

$$\begin{aligned} 2\sum_{n=1}^{\infty} a_n e(n\alpha) e^{-2\pi n y} - (1-(-1)^a)\Lambda(0) \\ &= \epsilon\sqrt{|d|}\sum_{n=1}^{\infty}\frac{a_n}{2\pi i}\int_{\Re(s)=7/4} n^{-s}\csc\left(\frac{\pi(s+a)}{2}\right)\big[|d|(y-i\alpha)\big]^{s-1}\,ds \\ &= \frac{2}{\pi}\epsilon i^{a+1}\sqrt{|d|}\sum_{n=1}^{\infty}\frac{a_n}{n}\frac{\left(\frac{|d|(\alpha+iy)}{n}\right)^a}{\frac{|d|(\alpha+iy)}{n} - \frac{n}{|d|(\alpha+iy)}}. \end{aligned}$$

If $\alpha|d|$ is not an integer then the last line is $O_\alpha(1)$ uniformly for $y \in (0,1)$, while if $\alpha|d| = n$ is an integer then we get $\frac{\epsilon i^a a_n}{\pi\sqrt{|d|}y} + O_\alpha(1)$. Since the left-hand side is periodic in α, we conclude that d is an integer and $a_n = a_{n+|d|}$, i.e. the coefficients a_n are periodic. Moreover, since $\Lambda(s)$ does not vanish for $\Re(s) > 1$, it follows from [28, Thm. 4] that there is a positive integer q dividing d and a primitive Dirichlet character $\chi \pmod q$ such that $L(s) = D(s)L(s,\chi)$, where $D(s) = \sum_{n \mid \frac{|d|}{q}} b_n n^{-s}$ for certain coefficients b_n, with $b_1 = 1$.

Let $\Lambda(s,\chi) = q^{s/2}\Gamma_{\mathbb{R}}(s+a')L(s,\chi)$ be the associated complete L-function. Then we have

$$\frac{\Lambda(s)}{\Lambda(s,\chi)} = \left(\frac{|d|}{q}\right)^{s/2}\frac{\Gamma_{\mathbb{R}}(s+a)}{\Gamma_{\mathbb{R}}(s+a')}D(s). \tag{26}$$

Moreover, it is easy to see that $D(s)\Lambda(s,\chi)$ does not vanish in some left half plane. Thus, to avoid concluding from (26) that $\Lambda(s)$ has poles at negative integers, it must be the case that and $a' = a$, so that $\chi(-1) = \operatorname{sgn} d$. From this and the functional equations for $\Lambda(s)$ and $\Lambda(s,\chi)$, it follows that $\frac{\Lambda(s,\chi)}{\Lambda(s,\overline{\chi})}\frac{D(s)}{D(1-s)}$ is an entire function. Note that for large $T > 0$, $D(s)/D(1-s)$ has $O(T)$ zeros and poles with imaginary part in $[-T,T]$. On the other hand, work of Fujii [9] shows that if χ_1 and χ_2 are distinct, primitive Dirichlet characters then $\Lambda(s,\chi_1)/\Lambda(s,\chi_2)$ has $\gg T\log T$ zeros and poles in that region. Thus, we must have $\chi = \overline{\chi}$, i.e. χ is quadratic and $q \operatorname{sgn} d$ is a fundamental discriminant.

Therefore, by (26) and the functional equations for $\Lambda(s)$ and $\Lambda(s,\chi)$, $D(s)$ satisfies the functional equation

$$D(s) = \epsilon \left(\frac{|d|}{q}\right)^{\frac{1}{2}-s} D(1-s). \tag{27}$$

Next, from the formula for $L(s)$, we have

$$\frac{D'}{D}(s) = \sum_{\substack{n\geq 2 \\ n\left|\left(\frac{|d|}{q}\right)^{\infty}\right.}} \left(\frac{\Lambda(n)\chi(n)}{\sqrt{n}} - c_n\right) n^{\frac{1}{2}-s},$$

where the notation $n\left|\left(\frac{|d|}{q}\right)^{\infty}\right.$ means that n is composed only of primes dividing $|d|/q$.

Now, from (27) and the estimate $\frac{\Lambda(n)\chi(n)}{\sqrt{n}} - c_n = O(n^{-\delta})$ it follows that $\frac{D'}{D}(s)$ is entire, and thus $D(s) = 1$ identically. Finally, invoking (27) once more, we have $|d| = q$ and $\epsilon = 1$. □

References

1. M. Ajtai, *The shortest vector problem in L_2 is NP-hard for randomized reductions (extended abstract)*, Proceeding STOC '98 Proceedings of the thirtieth annual ACM symposium on Theory of computing, 1998.
2. Gérard Ben Arous and Paul Bourgade, *Extreme gaps between eigenvalues of random matrices*, arXiv:1010.1294, 2011.
3. Estelle L. Basor and Torsten Ehrhardt, *Asymptotic formulas for determinants of a sum of finite Toeplitz and Hankel matrices*, Math. Nachr. **228** (2001), 5–45. MR 1845906 (2002d:47041)
4. Brian Conrey, *Notes on eigenvalue distributions for the classical compact groups*, Recent perspectives in random matrix theory and number theory, London Math. Soc. Lecture Note Ser., vol. 322, Cambridge Univ. Press, Cambridge, 2005, pp. 111–145. MR 2166460 (2006g:11177)
5. Edgar Costa and David Harvey, *Faster deterministic integer factorization*, arXiv:1201.2116, 2011.
6. Harold Davenport, *Multiplicative number theory*, third ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000, Revised and with a preface by Hugh L. Montgomery. MR 1790423 (2001f:11001)
7. P. Deift, A. Its, I. Krasovsky, and X. Zhou, *The Widom-Dyson constant for the gap probability in random matrix theory*, J. Comput. Appl. Math. **202** (2007), no. 1, 26–47. MR 2301810 (2008e:82027)
8. Percy Deift, Alexander Its, and Igor Krasovsky, *Asymptotics of Toeplitz, Hankel, and Toeplitz+Hankel determinants with Fisher-Hartwig singularities*, Ann. of Math. (2) **174** (2011), no. 2, 1243–1299. MR 2831118 (2012h:47063)
9. Akio Fujii, *On the zeros of Dirichlet L-functions. V*, Acta Arith. **28** (1975/76), no. 4, 395–403. MR 411182 (81g:10057a)
10. *GMP-ECM*, `http://ecm.gforge.inria.fr/`.
11. *GNU Linear Programming Kit*, `http://www.gnu.org/software/glpk/`.
12. Jerzy Kaczorowski and Alberto Perelli, *On the structure of the Selberg class. I. $0 \leq d \leq 1$*, Acta Math. **182** (1999), no. 2, 207–241. MR 1710182 (2000h:11097)
13. Nicholas M. Katz and Peter Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications, vol. 45, American Mathematical Society, Providence, RI, 1999. MR 1659828 (2000b:11070)
14. ———, *Zeroes of zeta functions and symmetry*, Bull. Amer. Math. Soc. (N.S.) **36** (1999), no. 1, 1–26. MR 1640151 (2000f:11114)
15. I. Krasikov, *Uniform bounds for Bessel functions*, J. Appl. Anal. **12** (2006), no. 1, 83–91. MR 2243854 (2008c:33002)

16. I. V. Krasovsky, *Gap probability in the spectrum of random matrices and asymptotics of polynomials orthogonal on an arc of the unit circle*, Int. Math. Res. Not. (2004), no. 25, 1249–1272. MR 2047176 (2005d:60086)
17. Youness Lamzouri, Xiannan Li, and Kannan Soundararajan, *The least quadratic non-residue, values of L-functions at $s = 1$, and related problems*, `arXiv:1309.3595` (2013).
18. A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, *Factoring polynomials with rational coefficients*, Math. Ann. **261** (1982), no. 4, 515–534. MR 682664 (84a:12002)
19. Robert Martin, *Bandlimited functions, curved manifolds, and self-adjoint extensions of symmetric operators*, ProQuest LLC, Ann Arbor, MI, 2008, Thesis (Ph.D.)–University of Waterloo (Canada). MR 2712567
20. Madan Lal Mehta, *Random matrices*, third ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004. MR 2129906 (2006b:82001)
21. Hugh L. Montgomery and Andrew M. Odlyzko, *Large deviations of sums of independent random variables*, Acta Arith. **49** (1988), no. 4, 427–434. MR 937937 (89m:11075)
22. A. M. Odlyzko, *On the distribution of spacings between zeros of the zeta function*, Math. Comp. **48** (1987), no. 177, 273–308. MR 866115 (88d:11082)
23. A. M. Odlyzko and H. J. J. te Riele, *Disproof of the Mertens conjecture*, J. Reine Angew. Math. **357** (1985), 138–160. MR 783538 (86m:11070)
24. Sami Omar, *Non-vanishing of Dirichlet L-functions at the central point*, Algorithmic number theory, Lecture Notes in Comput. Sci., vol. 5011, Springer, Berlin, 2008, pp. 443–453. MR 2467864 (2009k:11133)
25. J. M. Pollard, *Theorems on factorization and primality testing*, Proc. Cambridge Philos. Soc. **76** (1974), 521–528. MR 0354514 (50 #6992)
26. Michael Rubinstein, *Low-lying zeros of L-functions and random matrix theory*, Duke Math. J. **109** (2001), no. 1, 147–181. MR 1844208 (2002f:11114)
27. Michael Rubinstein and Peter Sarnak, *Chebyshev's bias*, Experiment. Math. **3** (1994), no. 3, 173–197. MR 1329368 (96d:11099)
28. Eric Saias and Andreas Weingartner, *Zeros of Dirichlet series with periodic coefficients*, Acta Arith. **140** (2009), no. 4, 335–344. MR 2570109 (2010m:11107)
29. K. Soundararajan, *Degree 1 elements of the Selberg class*, Expo. Math. **23** (2005), no. 1, 65–70. MR 2133337 (2006c:11104)
30. Volker Strassen, *Einige Resultate über Berechnungskomplexität*, Jber. Deutsch. Math.-Verein. **78** (1976/77), no. 1, 1–8. MR 0438807 (55 #11713)
31. M. N. S. Swamy, *Further properties of Morgan-Voyce polynomials*, Fibonacci Quart. **6** (1968), no. 2, 167–175. MR 0237470 (38 #5752)

A. R. B. and J. P. K.: School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, United Kingdom

G. A. H.: Department of Mathematics, The Ohio State University, 231 West 18th Ave, Columbus, OH 43210, United States